NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

INFLUENCING GAMEPLAY IN SUPPORT OF EARLY SYNTHETIC PROTOTYPING STUDIES

by

Douglas J. Ross

June 2016

Thesis Advisor: Rudolph Darken
Co-Advisor: Brian Vogt

This thesis was performed at the MOVES Institute

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2016	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE INFLUENCING GAMEPLAY IN SUPPORT OF EARLY SYNTHETIC PROTOTYPING STUDIES			**5. FUNDING NUMBERS**
6. AUTHOR(S) Douglas J. Ross			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number NPS.2016.0036-IR-EP7-A.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Early Synthetic Prototyping (ESP) is a concept being developed by the Army Capabilities Integration Center (ARCIC) to utilize a game environment and crowdsourcing techniques to receive end-user feedback on proposed acquisition programs early in the concept development stage. To be effective, ESP will need soldiers to participate, both to produce data and to interact with the game environment in such a way that the data is meaningful.

This study proposed a methodology for creating scoring algorithms and examined its ability to influence player behavior and enjoyment.

A group of students and faculty from the Naval Postgraduate School executed two scenarios in a VBS3 game environment. A scoring algorithm was applied to one scenario and data collected to determine the effect on player behavior and motivation.

The study found qualitative evidence that scoring mechanisms enhanced enjoyment and could influence desired behavior. However, quantitative data was not statistically significant to demonstrate a corresponding effect on gameplay. The results of this preliminary work can be used to support future studies on how to utilize scoring algorithms to support ESP research.

14. SUBJECT TERMS Early Synthetic Prototyping, acquisition, video games, crowdsourcing, Engineering Resilient Systems, science and technology, game environment			**15. NUMBER OF PAGES** 95
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

INFLUENCING GAMEPLAY IN SUPPORT OF EARLY SYNTHETIC PROTOTYPING STUDIES

Douglas J. Ross
Major, United States Army
B.S., United States Military Academy, 2002

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN
MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION (MOVES)**

from the

**NAVAL POSTGRADUATE SCHOOL
June 2016**

Approved by: Rudolph Darken
Thesis Advisor

Brian Vogt
Co-Advisor

Peter J. Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Early Synthetic Prototyping (ESP) is a concept being developed by the Army Capabilities Integration Center (ARCIC) to utilize a game environment and crowdsourcing techniques to receive end-user feedback on proposed acquisition programs early in the concept development stage. To be effective, ESP will need soldiers to participate, both to produce data and to interact with the game environment in such a way that the data is meaningful.

This study proposed a methodology for creating scoring algorithms and examined its ability to influence player behavior and enjoyment.

A group of students and faculty from the Naval Postgraduate School executed two scenarios in a VBS3 game environment. A scoring algorithm was applied to one scenario and data collected to determine the effect on player behavior and motivation.

The study found qualitative evidence that scoring mechanisms enhanced enjoyment and could influence desired behavior. However, quantitative data was not statistically significant to demonstrate a corresponding effect on gameplay. The results of this preliminary work can be used to support future studies on how to utilize scoring algorithms to support ESP research.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

3D	Three Dimensional
AAR	After Action Review
AI	Artificial Intelligence
AR	Army Regulation
ARCIC	Army Capabilities Integration Center
ASA(ALT)	Assistant Secretary of the Army for Acquisition, Logistics and Technology
CDD	Capabilities Development Document
COI	Critical Operational Issues
DOD	Department of Defense
DT	Developmental Testing
DT&E	Developmental Testing and Evaluation
DTT	Doctrine and Tactics Training
EMD	Engineering and Manufacturing Development
ESP	Early Synthetic Prototyping
GQM	Goal-Question-Metric Paradigm
ICD	Initial Capabilities Document
JCIDS	Joint Capabilities Integration and Development System
MMOG	Massively Multiplayer Online Game
MOE	Measure of Effectiveness
MOP	Measure of Performance
NET	New Equipment Training
OT	Operational Testing
OT&E	Operational Testing and Evaluation
SAM	Surface to Air Missile
STRAP	System Training Plan
T&E	Test and Evaluation
TEMP	Test and Evaluation Master Plan
TES	Test and Evaluation Strategy
TTP	Techniques, Tactics, and Procedures

USA	United States Army
USD[AT&L]	Undersecretary of Defense for Acquisition, Technology, and Logistics
USMC	United States Marine Corps
USN	United States Navy
VBS3	Virtual Battlespace 3

I. INTRODUCTION

A. PROBLEM STATEMENT

The U.S. Army has identified a lack of end user feedback early in the design process as a concern with current acquisition programs. Early synthetic prototyping (ESP) is a concept being developed by the Joint and Army Modeling and Simulation Division of the Army Capabilities Integration Center (ARCIC) to use a persistent game network and crowdsourcing techniques to explore design concepts to provide end user feedback early in the acquisition process (Vogt, 2015).

Initial studies have indicated that Soldiers would likely participate in ESP related studies as a means to influence the future force and that these studies can provide valuable insights to materiel developers. The military deputy to the assistant secretary of the Army for Acquisition, Logistics and Technology (ASA(ALT)) has expressed support for the program and ARCIC is currently developing requirements documents to gain funding for continued development of ESP systems for eventual integration into the acquisition process (Vogt, Megiveron, & Smith, 2015).

In order for ESP to generate the insights required to inform acquisition decisions, soldiers must participate in studies that generate useful data. Also, soldiers must act in a tactically sound manner to ensure that data collected is accurate and useful. Since soldiers will be untrained on the prototype equipment they are provided in the game environment, they may be unaware of how to utilize unique capabilities of prototype systems to enhance mission effectiveness. It may be necessary to design the game environment in a manner that influences players to behave in a manner that allows them to realize the benefits of prototype systems in order to ensure that data collected during studies relevant to research questions that need answered.

Since ESP is designed where soldiers can access the game environment at their convenience, observers will not be able to interact with players during gameplay. These observers would be able to interact with players and discuss how they might best utilize the capabilities of the prototype systems they are provided. To offset the lack of controllers, it may be necessary to design the game environment to provide a mechanism to influence players to utilize the unique capabilities of the prototype systems.

"Gamification" is a concept of applying game mechanics to human activity to promote engagement, enjoyment, and motivation. Commonly, gamification has been applied activities that are traditionally not games such as: (1) physical workouts to encourage people to stick with a physical exercise plan, and (2) customer rewards to encourage consumer loyalty. Gamification usually involves player rewards such as badges, "rank," or scoring (Fitz-Walter).

A potential method for influencing player behavior within ESP is to apply a gamification scoring algorithm to gameplay that rewards players who utilize the unique capabilities provided by the prototype systems without specifically telling them how to use the prototype system. This thesis investigates this specific issue.

B. RESEARCH QUESTIONS

This study focused on two research questions to gain insight on how the use of scoring algorithms in an ESP environment will affect player behavior and experience.

1. Do changes in a scoring algorithm affect player behavior?
2. Do changes in a scoring algorithm affect player enjoyment?

Exploratory question:

How can we design the ESP game environment to ensure that soldiers generate meaningful data without decreasing enjoyment?

C. SCOPE OF THIS THESIS

The objective of this thesis is to propose a methodology for applying elements of gamification to influence player behavior in a game environment to ensure that data collected from gameplay provide accurate, relevant data to answer engineering design questions used to inform acquisition decisions. This study examined how to derive game metrics to collect to inform engineering design questions. It further discusses how to determine what player behaviors are required to generate desired game metrics and how to develop measures of performance to evaluate a player's performance with respect to designated game metrics. This thesis then proposes a method for combining measures of effectiveness related to tactical mission success and measures of performance related to required game metrics to derive scoring algorithms that support acquisition studies.

D. BENEFITS OF STUDY

This study supports the U.S. Army development of an online, virtual crowdsourcing environment by providing ESP developers with increased awareness of the effects of scoring mechanisms on player behavior, experience, and motivation. This knowledge will enable researchers to develop the game environment and scenarios in a manner that allows for player enjoyment while providing the necessary data to answer acquisition program information requirements.

E. THESIS ORGANIZATION

This thesis is organized as follows. Chapter I provides the problem statement, research questions, thesis scope, and benefits of the study. Chapter II provides background information on guidance and requirements leading to the development of ESP, studies in crowdsourcing systems, ESP methodology and challenges, gamification techniques, and studies on game analytics and building the online gaming community. Chapter III discusses how metrics are developed in the DOD acquisition test and evaluation process and provides a methodology for

developing metrics for use in ESP studies. Chapter IV discusses experimental design, participant metrics, and methods for conducting the study. Chapter V discusses the results of the study. Chapter VI provides lessons learned, future research requirements, and a conclusion.

II. BACKGROUND

A. EARLY SYNTHETIC PROTOTYPING

The United States military has a persistent need for innovation to maintain competitive advantage on the modern day battlefield. The process of designing, developing, and fielding new military equipment remains a complex and time-consuming process and when these design efforts fail, costs are typically very high (McGroarty, 2015).

A contributing factor in many failed acquisition programs is that there is no mechanism for communication between engineers and soldiers who will utilize new equipment prior to production of physical prototypes. This lack of communication, combined with different terminology used by the engineering and military communities results in confusion about what needs to be designed and products are often produced that do not meet the military's requirements (McGroarty, 2015).

Another issue is a lack of testing early in the design process. Evaluators frequently do not conduct testing until product developers have produced a working prototype. This results in design deficiencies being identified late in the acquisition process when cost to correct deficiencies becomes much greater. The U.S. Army Capabilities Integration Center (ARCIC) is developing a process called Early Synthetic Prototyping (ESP) to address these common causes for acquisition failures (McGroarty, 2015).

B. WHY ESP – GUIDANCE AND REQUIREMENT

The secretary of defense, the Honorable Chuck Hagel, stated, "A world where our military lacks a decisive edge would be less stable [and] less secure for both the United States and our Allies" (Parker, 2014). The U.S. Military has long been able to employ superior technology to gain decisive advantage against our adversaries. However, the Undersecretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]), the Honorable Frank Kendall, has warned, "our technological superiority is very much at risk" (Freedberg, 2014). Many modern

day adversaries, unencumbered by bureaucratic requirements, are able to utilize nimble design processes that provide a rapid innovation trajectory despite a disadvantage in available resources (Murray, 2014).

Contrasting to the nimble design and innovation processes employed by some of our adversaries, soldiers provide feedback that the process to design and field new technology to address operational requirements is too long and fielded equipment is often inferior to products that can be procured commercially-off-the-shelf for less money. Additionally, constricting budgets following over a decade of sustained conflict make acquisition failures even more costly (Murray, 2014).

C. CROWDSOURCING – A PREDECESSOR TO ESP

ESP will build upon recent commercial successes of crowdsourcing systems. In the June 2006 issue of Wired Magazine, Jeff Howe defines crowdsourcing using the following definition: “Simply defined, crowdsourcing represents the act of a company or institution taming a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call” (Howe, 2006).

Crowdsourcing is developing in part due to the diversity of the marketplace. No matter how diverse and large a company tries to make their design teams, they are typically poor representations of the crowds in the marketplace they are designing for. For this reason, design teams and crowds typically reach conclusions in two distinct manners. Design teams typically rely on experts and tend to be hierarchical in nature. Conversely, in a crowd, individuals do not possess rank. This allows the marketplace to benefit from including non-experts and amateurs (Brabham, 2008).

In the book *The Wisdom of Crowds*, James Surowiecki states that crowds have the ability to develop an intelligence that is greater than the smartest people in the crowd. This happens because crowds are able aggregate solutions. Surowiecki states, “With most things, the average is mediocrity. With decision

making, it's often excellence. You could say it's as if we've been programmed to be collectively smart" (Surowiecki, 2005).

Doan describes crowdsourcing systems as being designed to utilize this collective wisdom of crowds of consumers to solve problems posed by an interested party, typically commercial corporations. Crowdsourcing differs from open source systems in that solutions derived from crowdsourcing are the property of the corporation that solicited feedback from the crowd and the owner of the crowdsourced solutions is free to profit from insights gained through crowdsourcing (Doan and Halevy, 2011).

Research from Poetz and Schreier demonstrated that the insights gained from crowdsourcing are quite valuable. Bamed/MAM group, an Austria-based company that manufactures baby products participated in study where ideas for new products from the company's internal design team were compared with user-generated ideas received after an open call placed on the company's website and Internet forums. Executives from the company then conducted a blind evaluation of the expert and user generated innovation ideas. The study found that customer-generated ideas were generally superior in terms of novelty and value to the consumer than expert generated idea. However, expert-generated ideas were typically considered more feasible. Overall, three user generated ideas were identified as performing well in all three categories compared to only one idea generated from the company's internal design team. (Poetz and Schreier, 2012)

D. A METHODOLOGY TO INCREASE DESIGN EFFICIENCY AND RESPONSIVENESS

ESP seeks to leverage lessons learned from crowdsourcing systems by integrating the end user, soldiers, into the design process during initial acquisition planning stages in order to alleviate issues stemming from miscommunication and lack of early testing. The intent is to create a means to communicate between the engineers who will design and build new systems and the soldiers that will employ them (McGroarty, 2014). Figure 1 provides a visual depiction of the ESP concept.

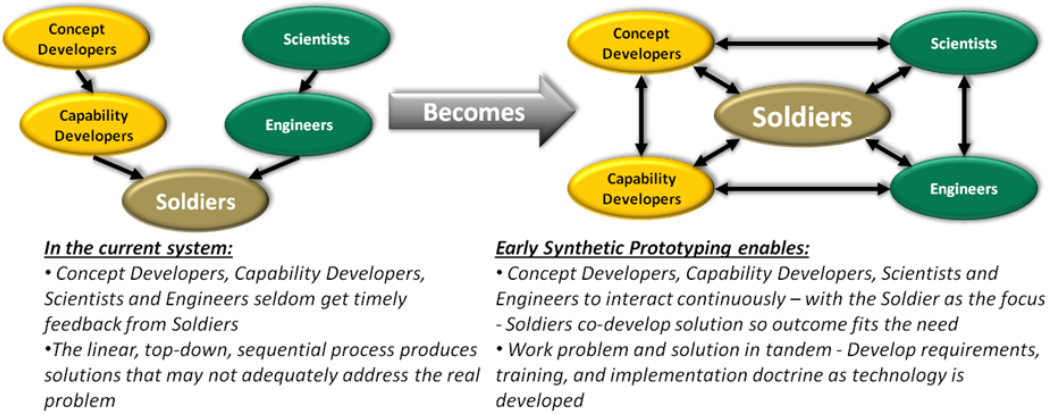


Figure 1. ESP Integrates Soldiers into the Design Process. Source: Vogt (2014).

Vogt, Megiveron, and Smith describe ESP as currently in the prototype stage. ARCIC is using the working ESP Schema to understand system requirements that will facilitate creativity and enable innovation using ESP. The current vision is that ESP will enable soldiers to assess future concepts and capabilities in a persistent game environment that will be available both on and off duty. Soldier feedback from the game environment will then be used to inform system design and material solution research. Feedback will also be used to examine force organization and doctrine to most effectively employ new systems. (Vogt, Megiveron, & Smith, 2015)

The process will begin when concept and capability developers and engineers propose doctrine, organization, or material solutions to identified warfighting requirements. These solutions will be modeled in a game environment and scenarios created that will enable researchers to conduct studies to answer identified information requirements. Soldiers across the Army will then be able to access these scenarios in a persistent on-line game environment. Soldiers will be provided information on proposed acquisition systems and will be provided the opportunity to make modifications prior to playing the scenarios. After playing the scenarios, soldiers will have the opportunity to provide feedback and

recommendations regarding the prototype being tested. In addition to qualitative soldier feedback, the system will be able to collect quantitative metrics related to system performance. By integrating soldiers into the design process, the Army envisions that they will be able to produce and explore orders of magnitudes more design alternatives than current acquisition methods allow. (Vogt, Megiveron, & Smith, 2015)

E. CHALLENGES TO SUCCESS

McGroarty stated that ESP analytic requirements differ from existing commercial game and simulation engines. Most games collect traditional metrics that provide data to facilitate the implementation of scoring algorithms. In order to answer acquisition related research questions, ESP must be able to collect a new class of metrics that focus on the requirements and desires of the user. ESP must be able to determine, not only what a player did, but also provide insight on how and why they took the actions that they did. The system must also be able to assess subjective metrics such as frustration and sources of frustration. (McGroarty, 2014)

A primary requirement for ESP to be successful is Soldier participation. If soldiers do not play the game, ESP will be unable to collect the required metrics to inform engineering and acquisition decisions. This requires knowledge of what type of games soldiers typically play and an understanding of what would motivate soldiers to participate in ESP related studies in their free time. This understanding will need to be combined with continuous assessments of soldier perception of the game environment to ensure that new scenarios are developed in a manner that will enable the environment to evolve in a manner that will encourage continued participation by soldiers and the development of a loyal online gaming community. (Vogt, Megiveron, & Smith, 2015)

Additionally, the feedback received from the system must be valuable. The qualitative and quantitative data generated by the system must address research questions in a manner that is valuable to concept and capability developers. This

will require developing scenarios and mechanisms to encourage soldiers to utilize prototypes in a manner that will generate meaningful data. In addition, questionnaires, surveys, and other subjective data collection methods must be designed to address specific research questions for each acquisition program. (Vogt, Megiveron, & Smith, 2015)

Soldiers will lack training on equipment being evaluated in the ESP environment. AR 350-1 directs the Army utilize a system of systems approach when fielding new equipment to modernize units. A key aspect of this approach is proving units with New Equipment Training (NET) to train the unit how to utilize the new systems. As part of the product development cycle, system training plans (STRAP) are documented in the Joint Capabilities Integration and Development System (JCIDS) capability requirements documents beginning with the Initial Capability Document (ICD). (U.S. Army, 2014a)

For systems that are likely to affect a change in the way a unit fights, which will likely include a large proportion of systems that are evaluated in the ESP environment; NET training is composed to three components. Operator NET trains soldiers on the capabilities and operation of the new system. Maintenance NET trains units on the upkeep and maintenance of the new equipment. Unit leaders will receive doctrine and tactics training (DTT) that will train them on proper tactical use of new equipment. Operator NET and DTT are critical components for units learning how to integrate newly fielded equipment into unit operations. (U.S. Army, 2014b)

Soldiers participating in ESP studies will not have the benefit of NET. This may affect their ability to successfully incorporate prototype equipment into operations in the ESP environment. Failure to utilize the unique capabilities provided by prototype equipment is likely to result in mission failure, which participants may inaccurately attribute to poor equipment. It is unlikely that soldiers will be willing to receive NET in the game environment before executing an ESP study. Therefore, thought must be given on how to design the game environment in a manner that encourages study participants to utilize prototype systems in a

manner that makes tactical sense and takes advantage of new capabilities provided by the equipment.

Researchers also must be careful not to design the game environment in a manner that restricts creativity. While prototype systems are designed to provide a specific capability, the tactics discussing how best to employ the system will not be developed yet. The game environment needs to encourage players to utilize the capabilities provided by the prototypes, while allowing them the freedom to determine how best to employ the systems.

F. GAMIFICATION – EFFECTS OF POINTS AND MEANING ON USER MOTIVATION

Meckler et al. define gamification as "the use of game design elements (e.g., points, leaderboards and badges) in non-game contexts, to promote user engagement" (p. 1138). Their research discusses potential benefits of elements of gamification, particularly scoring systems and meaningful framing, for improving the motivation and performance of individuals conducting tasks. Many of the lessons learned from their study are applicable to motivating soldiers to participate in ESP studies (Meckler, Brühlmann, Opwis, and Tuch, 2013).

In Disassembling Gamification, Meckler describes meaningful framing as "acknowledging a participants' contribution to a scientific cause." (pg 1138). Framing an action is accomplished by providing a purpose for a task that is deemed valuable by the person tasked to accomplish it. Their research indicated that framing could provide a form of intrinsic motivation reward caused by an inherent desire in people to contribute to improving the world around them (Meckler et al., 2013). In a related study, researchers determined they were able to increase the likelihood of participation in an image tagging task if participants were informed that their efforts would be used in efforts to identify tumor cells (Chandler and Kapelner, 2010). Researchers from Nanyang Technological University, Singapore conducted a study suggesting that participants preferred a "gamified" version of an

image tagging task, although their motivation did not translate to better quality of performance (Goh and Lee, 2011).

An ARCIC operational test conducted with the Brigade Modernization Command in December 2014 supports the hypothesis that meaningful framing contributes to motivation. During this study, soldiers were given an overview of ESP prior to conducting a brief study utilizing the Virtual Battlespace 3 (VBS3) game environment. In a survey conducted after the study, 77% of soldiers felt that the study had been an effective use of their time (55% very effective, 22% semi-effective). In addition, 80% indicated that they would be likely (65%) or somewhat likely (15%) to participate in future studies if they contributed to shaping the future force. One participant indicated that, even though he did not play video games for entertainment, he would find time to participate in ESP (Vogt, Megiveron, & Smith, 2015).

Meckler's experiment indicated that both scoring and framing provided effects that would be beneficial to ESP studies. The presence of a scoring algorithm provided motivation to perform the task and participants produced significantly more tags than those who were not awarded a score for their performance. Framing on the other hand resulted in tags of a higher quality compared to participants who were not provided a purpose for their task. The presence of both scoring mechanisms and meaning framing were shown to have a positive impact on intrinsic motivation of participants. The study also indicated a positive interactive effect on motivation when scoring mechanisms and framing were both presented to the participant (Meckler et al., 2013).

In Game Reward Systems, Wang and Sun describe how gaming reward systems contribute to enhancing user experience. Their study examines how reward systems vary based the type of game and how their effects vary based on players' individual preference and motivations. They describe eight forms of reward used in video games: 1) score systems, 2) experience points, 3) item granting, 4) collectible resources, 5) achievement systems, 6) feedback messages, 7) plot animations and pictures, and 8) unlocking mechanisms and

provide information on how players utilize these rewards, the social benefits a player enjoys when receiving rewards, and how reward systems contribute to player enjoyment (Wang and Sun, 2011).

G. BUILDING AND ONLINE COMMUNITY

Chapters 28–30 of El-Nasr et al's *Game Analytics* text discuss the use of analytics and player communities in massively multiplayer online games (MMOG). These chapters are concerned with understanding player behavior in MMOGs and the social dimensions that add new depth to the actions available to a player. The book also explores how social architecture in the game environments is designed to encourage collaboration and maximize the opportunities for players to interact (El-Nasr et al., 2013).

As DOD developed game environments are unlikely to provide entertainment value equivalent to commercially developed games, understanding the motivation provided by participation in online communities is beneficial to developing a core of ESP participants. From an individual perspective, player experience in MMOGs and single-player games would be very similar if the interactive component was removed. Indeed, most activities that players conduct in MMOGs are also available in single-player games. What makes MMOGs unique is how their social architecture facilitates social interaction and how repeated player interactions develop a dedicated community that influences player to return to the multiplayer environment. As games, MMOGs are not superior to single-player games. It is the interaction with other players that provides their attractiveness to gamers (El-Nasr et al., 2013).

The current crop of the most popular MMOGs, those with 100,000(+) subscribers, typically follows a similar formula for developing their gaming communities. New players begin as level 1 characters with minimal abilities and equipment. Players gain levels and attributes through the completion of quests or missions in the environment. Missions are initially simple and do not require cooperation between players to complete. However, as players gain experience

and power, the missions become more complex. In order to continue advancing their characters, players must interact with other players and form alliances to complete these more complex tasks. As players are forced to cooperate to complete these more complex tasks, they begin to value the relationships they form in the game environment. These valued relationships are what MMOGs rely upon to maintain their popularity in today's competitive gaming market. The military can leverage this concept in an ESP environment where players will share an identity as military members. This shared identity will likely allow relationships to develop quicker and result in relationships that players value more than relationships built in a commercial MMOG environment (El-Nasr et al., 2013).

H. GAME ANALYTICS

In order to provide useful feedback to product designers, ESP systems must be able to provide insight into how prototypes perform in the game environment. Game analytics is a developing field that provides the type of information that ESP researchers require. Game analytics has gained importance due to the increased competition that has developed in the game industry as technology has enabled an increasing number of companies to introduce games into an already competitive marketplace. Analytics gain their influence from data-driven business intelligence practices. Companies frequently focus analytic efforts on understanding their users, specifically focusing on what motivates them to purchase and play games and the experiences players gain from interacting with their products (El-Nasr et al., 2013).

El-Nasr discusses a variety of techniques for establishing game telemetry systems. The text defines telemetry systems as any technology that enables the collection and measurement of game data remotely over a distance. Telemetry systems are typically found in all on-line games today and allow researchers to move beyond traditional focus groups, beta-tests, and surveys and continuously collect data on how actual customers interact with a game and how interactions change over time. These types of systems will be necessary for collecting useful

data from ESP related studies that are intended to be available for soldiers to access on-line (El-Nasr et al., 2013).

THIS PAGE INTENTIONALLY LEFT BLANK

III. METRIC DEVELOPMENT

A. MEASURES AND METRICS

Developing a metric system for test and evaluation (T&E) is a key component of the acquisition process. Metrics are measurements that enable program managers to determine if an acquisition program is making progress towards meeting the system requirements outlined in the JCIDS process. An effective metrics program evaluates a program against organizational goals to support project management decisions utilizing a 3-step cycle that includes developing a metrics plan, implementing the plan, and evaluating the metrics program (Perkins, Peterson and Smith, 2003). The metrics program cycle as described by Perkins, Peterson, and Smith is shown in Figure 2.

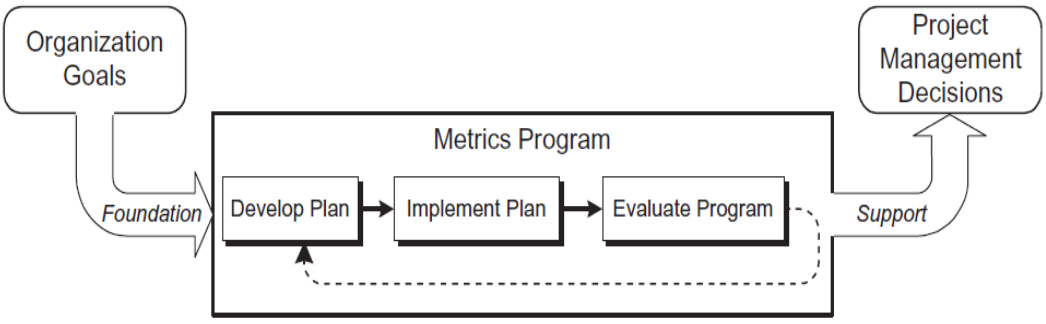


Figure 2. Metrics Program Cycle. Source Perkins, Peterson, and Smith (2003).

The goal-question-metric (GQM) paradigm is frequently used to support developing a metrics program plan. GQM is based on five key concepts:

1. Processes have associated goals.

2. Each goal has one or more associated questions related to its accomplishment.

3. One or more metrics are required to answer each question.

4. Each metric requires two or more measurements to determine progress.

5. Measurements provide data to produce the metric.

The process begins by declaring well-defined, validated goals that are worded in a manner to make them measurable and verifiable. Each goal is then examined to derive questions that describe how progress will be measured and metrics are developed to answer the research questions. Once a list of metrics is developed, program developers select measurements that will determine progress of each metric and determine how the measurements will be evaluated to produce the required metrics (Perkins, Peterson, and Smith, 2003). The Goal, Question, Metric Paradigm is shown in Figure 3.

Figure 3. The Goal, Question, Metric Paradigm. Source Perkins, Peterson, and Smith (2003).

B. METRIC DEVELOPMENT IN DOD ACQUISITIONS

Metric development of test and evaluation (T&E) in DOD acquisitions begins with the production of a Test and Evaluation Strategy (TES). The TES is an early planning document that provides an initial overview of T&E activities for the entire acquisition cycle from tech development through fielding of the validated end product. The TES is the primary T&E planning document used during the technology development phase of an acquisitions program. The TES describes risk reduction efforts to include developmental test and evaluation (DT&E) and operational test and evaluation (OT&E) that will be used to evaluate the operational effectiveness of a prototype system prior to fielding to operational units (Department of Defense, 2009).

As the program matures, the TES is refined into a more detailed Test and Evaluation Master Plan (TEMP) when the program is ready to enter into the EMD

phase of the acquisition process. The TEMP provides the overall structure for the T&E program and relates the test management strategy to Critical Operational Issues (COIs) that are documented in the Capabilities Development Document (CDD) that describes the functional specifications a prototype system must perform to in order to be considered ready for fielding (Department of Defense, 2009).

The TEM will include a combination of Measures of Effectiveness (MOE) and Measures of Performance (MOP). In acquisition T&E, MOEs describe how well a unit equipped with a prototype system is able to accomplish mission objectives and desired results. MOEs are directly related to COIs that the prototype system is designed to address. MOEs are decomposed into MOPs that evaluate a prototype's performance against quantifiable Key Performance Parameters (KPP). The figure below depicts the traceability of COIs through MOEs to MOPs during T&E (Department of Defense, 2005). Figure 4 depicts the traceability of MOP and MOE metrics to COIs.

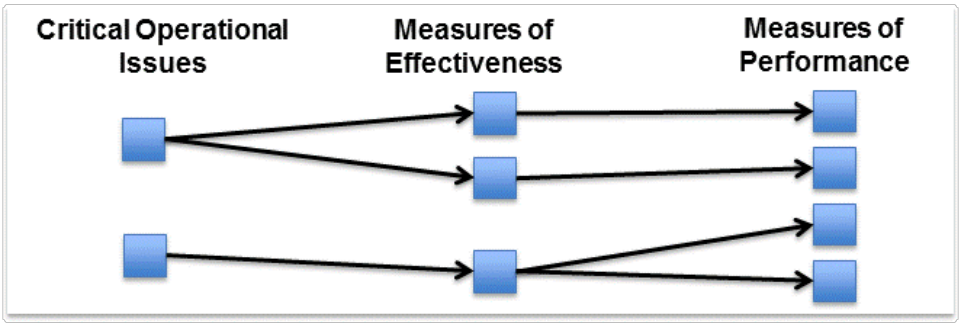


Figure 4. MOE/MOP Traceability Diagram

Developmental testing (DT) is conducted by the contractor to ensure that prototypes are meeting the specification required by the CDD. DT is primarily conducted during the technology maturation and EMD phases of the acquisition program. The purpose of the DT&E program is to ensure that prototypes meet the specifications required to meet key performance parameters (KPP) in order to satisfy the COIs laid out in the Initial Capabilities Document (ICD). The TEMP

provides KPPs that they prototype must satisfy prior to moving into OT. These KPPs are equivalent to the goals utilized in the GQM paradigm discussed previously. The KPPs outlined in the TEMP are then used to derive MOPs that measure the performance of the prototype system during DT (Department of Defense 2015).

Operational testing (OT) is conducted by the Department of Defense at the beginning of the production and deployment phase once contractors have a working prototype that has successfully met KPPs tested during the EMD phase. OT is done by operational units that are equipped with prototype systems. Evaluation during this phase is not focused on system performance, but rather how well a unit performs when equipped with the prototype system. The OT utilized COIs to determine operational goals for OT&E. These goals are then utilized to derive MOEs that units will be evaluated against. As the goal of acquisition programs are to field systems that enhance the operational effectiveness of military units, OT&E is the major hurdle that an acquisition program must clear before it is deemed ready for fielding to the force (Department of Defense, 2015).

C. A METHODOLOGY FOR DEVELOPING GAME METRICS FOR ESP STUDIES

The metric development methods described by the GQM paradigm and used in DOD acquisitions can be applied to develop scoring algorithms to support ESP studies. Scoring in the ESP environment can serve two purposes; to increase player enjoyment and their likelihood to participate in future studies and to influence behavior to increase the value of data collected. When using scoring algorithms to influence player behavior, the goal is to encourage players to utilize prototype capabilities in a sound tactical manner that will contribute to mission accomplishment.

Similar to metric development in DOD acquisitions, ESP metrics should maintain traceability to the research questions that the study is seeking to address. This ensures that the scoring mechanisms are encouraging the intended behaviors

and do not actually decrease the value of the data being collected. Where acquisition metrics trace back to COIs, ESP metrics should trace to specific research questions the study is seeking to address. Scenario developers will determine what metric data is required to answer the research questions and will design the scenario in a manner that allows the gameplay to support collecting the required data.

Once a scenario is developed, the study team will determine what outcomes would constitute mission success. These outcomes are provided to the player as game objectives. Game objectives can be either rewarding or punitive in nature. Scoring algorithms will provide flat rate scores to successfully meeting rewarding mission objectives or reduce scores by a flat rate for violating punitive mission objectives. The value of completing an objective will be proportional to its overall significance in contributing to mission success. Similar to MOEs in acquisition T&E, accomplishing the game objectives is the primary focus of determining the overall evaluation of player performance. This maintains a focus on sound tactical performance focused on mission success.

Once game objectives are identified, the study team will determine how unique prototype capabilities contribute to accomplishing specific objectives. Metrics will then be developed to measure how well a player utilizes the capabilities provided to them, similar to MOPs in acquisition metrics. A scaled score will be included in the scoring algorithm to evaluate how well a player utilizes specific capabilities. Researchers will need to determine the proper amount of weight to apply to performance measures for each study. Enough weight must be applied to provide an incentive to use prototype capabilities. However, care should be taken that too much weight is applied to performance metrics so that a player prioritizes utilizing the prototype over accomplishing the mission. Placing too much weight on prototype performance may also restrict a player's desire to investigate new tactics, techniques, and procedures (TTP) that may prove superior to contributing to mission accomplishment than TTPs envisioned by the research team.

Once mission objectives and performance metrics are identified, the research team needs to verify that all metrics trace back to a research question. This will ensure that all elements of the scoring algorithm contribute to mission success and will prevent the inclusion of secondary objectives. Secondary objectives are game objectives that are presented to a player which have no impact on accomplishing the primary objective of a scenario. Studies have shown that including secondary game objectives has the potential to decrease a player's motivation to participate in future studies. Also, game objectives that do not contribute to mission success will likely influence players to utilize improper tactics and my decrease the quality of data being collected (Anderson, Liu, Snider, Szeto, Cooper, and Popović, 2011).

Weighting evaluation criteria is the final step in finalizing the scoring algorithm. As discussed previously, scoring mechanisms should prioritize sound tactics and mission success over any specific TTPs on prototype use. Therefore, a weighting where 2/3 of a player's score is generated by accomplishing mission objectives and 1/3 is received from utilizing prototype capabilities is recommended. This weighting will need to be evaluated for each study and may be altered if it is determined to provide either too much or too little influence on player behavior in the game environment.

IV. METHODS

A. PARTICIPANTS

This study utilized participants from the students and faculty of the Naval Postgraduate School (NPS). The NPS Institutional Review Board approved the study, protocol number NPS.2016.0036-IR-EP7-A. All participants were either current or former members of the either the U.S. or partnered military forces. Recruitment utilized a combination of email, flyers, social engagement, and the NPS muster page to solicit volunteers. The study attempted to focus recruiting efforts on the segment of the NPS population who regularly play video games in their free time. This focus was because the majority of the soldiers that eventually participate in ESP studies will be drawn from a population that regularly plays video games.

Twenty participants completed the experiment, including 15 active duty U.S. military (4x USA, 6x USMC, 5x USN), four partnered military, and one prior service member of the USN. Although the study attempted to recruit volunteers who played video games on a regular basis, the majority of the respondents were not in the target audience. Only three volunteers stated that they played games for more than ten hours a week, compared to ten volunteers who stated they did not play games at all. The remaining seven volunteers played between 2 and 3 hours per week. Tables 1 and 2 depict the participant demographic information:

Table 1. Demographic Statistics of Study Participants

DEMOGRAPHICS (Numerical Data)	Scnario 1 Scored (SD)	Scenario 2 Scored (SD)	Total (SD)
Age	33.40 (4.12)	38.40 (5.85)	35.90 (5.55)
Years of Service	12.50 (4.95)	13.89 (4.88)	13.00 (4.76
Number Combat Tours	1.90 (1.91)	1.50 (1.58)	1.70 (1.72)
Weekly Video Game Hours	3.20 (6.41)	3.70 (6.50)	3.45 (6.29)
Weekly Military Themed Game Hours	0.55 (1.07)	1.05 (3.15)	0.80 (2.30)

Table 2. Categorical Data of Study Participants

DEMOGRAPHICS (Catagorical Data)	Scnario 1 Scored	Scenario 2 Scored	Total Count
Military Status			
Active Duty USA	1	3	4
Active Duty USMC	3	3	6
Active Duty USN	3	2	5
Prior Service USN	0	1	1
Active Duty Partnered Miitary Service	3	1	4
Sex			
Male	9	10	19
Female	1	0	1
Have Deployed			
Yes	7	6	13
No	3	4	7

B. DESIGN

The study utilized the Virtual Battlespace 3 (VBS3) game environment and was conducted in five phases: introduction and demographic survey, VBS3 training, execute unscored scenario, execute scored scenario, and post-task survey.

VBS3 was developed by Bohemia Interactive and is the U.S. Army's primary game environment used in the Games for Training program (Bohemia Interactive, 2016). VBS3 utilizes a 3D, first-person shooter design to generate realistic, semi-immersive environments. VBS3 provides developers with a large catalogue of U.S., foreign military, and civilian personnel and equipment, as well as geo-typical and geo-specific terrain for a number of training areas. The game environment adds realism by incorporating realistic capabilities and physical characteristics for its personnel and equipment models (Milgaming, 2016). The program also provides a built-in AAR feature that can record game play and provide detailed accounting of all engagements that occur during a scenario (Bohemia Interactive, 2016). VBS3 is unlikely to be the game environment selected to support for ESP studies, however the realism provided by the game is a good representation of the type of game environment the system will seek to provide to the user.

Two single-player scenarios were developed for the study. The first was a dismounted raid scenario. The second scenario was a mounted hostage rescue mission.

1. Controlling Variability

The primary sources of variability in this study included tactical actions of the participant and the performance of artificial intelligence (AI).

The study attempted to control unwanted variability from study participants utilizing a number of methods. The unscored scenario was always conducted first in order to prevent the user being influenced by the knowledge that scoring algorithms existed for the scenarios. Conducting the scored scenario first could have caused unintentional influence on player behavior in the unscored scenario. Other methods of control included using a single study member to conduct all iterations of the study including: set-up, in-brief, scenario briefs, training, and administering of post-task survey. Other methods of controlling variability were requiring all study participants to execute a standard training scenario to provide a common baseline understanding of game controls and providing standardized briefings for each scenario to ensure that participants were provided with the same information in preparation for scenario execution.

2. Randomization

Participants were randomly assigned to two treatment groups. The first treatment group applied a scoring algorithm to the dismounted raid scenario, while the second treatment group applied a scoring algorithm to the mounted hostage rescue mission.

3. Replication

Each scenario began with a standard starting position for friendly and enemy forces. Enemy forces were initially static, providing the same initial starting conditions upon first contact between enemy and friendly forces. The study team

maintained the VBS3 scenario files that could allow the game environment to be replicated for future studies.

C. SCENARIOS

1. Training Scenario

The initial training scenario was built using the VBS3 Twentynine Palms map. For the training scenario, participants were equipped with an Australian M4A5 SD Elcan assault rifle with eight 30-round magazines of 5.56 mm ammunition. The rifle is suppressed, has a maximum effective range of 500 meters, and can fire in semi- or fully-automatic modes. The M4A5 rifle is depicted in Figure 5. Participants were also provided a suppressed M9 Berretta pistol with three 15-round magazines of 9 mm ammunition, binoculars, two M67 fragmentation grenades, two white smoke grenades, and two M183 satchel charges.

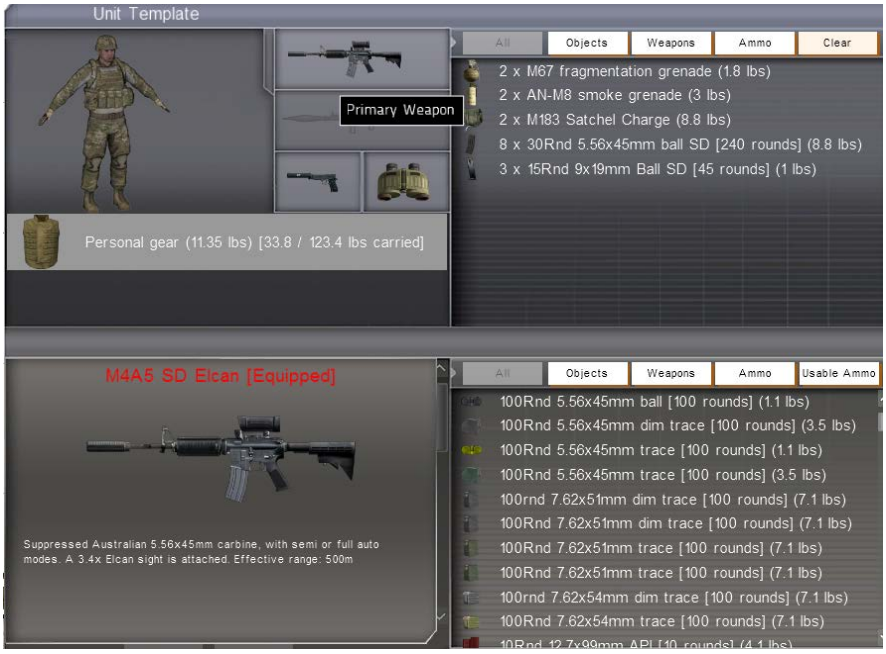


Figure 5. Soldier Basic Load for Training Scenario

As part of the training scenario, participants were trained on basic movement controls to include crawling, walking, running, jumping, leaning left and right, and changing body position between standing, crouching, and prone positions. Participants practiced marksmanship by engaging four target silhouettes and were instructed on use of fragmentation and smoke grenades. Once participants stated and demonstrated that they were proficient at controlling their avatar during dismounted operations, they transitioned to the mounted portion of the training.

During mounted training, participants were instructed on how to mount, dismount, and drive a USMC LAV25A2 wheeled armored troop transport that is shown in Figure 6. Participants were instructed to mount the vehicle and were instructed to navigate along a paved road through a wire obstacle to a house that was identified by the study controller. Participants practiced maneuvering the vehicle until they stated they were comfortable with controls and navigation.

Figure 6. USMC LAV25A2

Once participants were proficient with maneuvering the LAV, they were instructed to dismount near a one-story house to train on capturing and interacting with a civilian hostage. Participants were shown how to open and close doors in order to enter a building and how a proximity trigger would attach the hostage to

their avatar allowing them to provide commands to the hostage avatar. Participants were given instruction on how to provide commands to the hostage including movement commands and commanding the hostage avatar to mount the LAV. Training was completed by instructing participants on how to place and detonate the M183 satchel charge and a demonstration of its use and blast radius. Once training was complete, study participants were provided an opportunity to practice any skills that they desired prior to executing the first test scenario.

2. Test Scenario 1 – Dismounted Raid

For the dismounted test scenario, participants were instructed that they were being equipped with a prototype rifle with a maximum effective range of 500 meters. Participants were briefed that the military was interested in studying the effectiveness of units that were provided with a primary weapon that provided increased maximum effective range with no accompanying degradation in the weapon's performance in a close quarter environment. Participants were equipped with a suppressed Australian M4A5 SD Elcan assault rifle that served as the prototype rifle system for the study. The scenario brief provided to participants is provided in Appendix C.

The study team utilized the metric development methodology discussed in Chapter III and determined that a dismounted raid scenario utilizing a combination of open terrain and a small village would be best suited for studying how participants would utilize the new capability. The scenario was initially developed using military operational graphics on paper. The team then identified terrain on the VBS3 Sarhani map that would meet the scenario's objectives.

In order to provide participants with a tactical decision that would be identifiable to the study team, the team designed the scenario with two distinct avenues of approach to the objective village. The first approach incorporated a combination of vegetation and terrain masking to provide a route to the edge of the village that provided excellent cover and concealment, but also offered poor observation and fields of fire that would restrict the ability to use the enhanced

range afforded by the prototype rifle. Figure 7 shows an image of the northern, concealed route to the objective.

Figure 7. Northern Avenue of Approach – Excellent Cover and Concealment

A second avenue of approach utilized elevated terrain that gradually sloped towards the village. This approach provided only sporadic cover and concealment for the participants as they approached the village. However, this avenue of approach offered good observation and fields of fire that would enable the user to utilize the superior range of the prototype system. The southern, open fields of fire route is shown in Figure 8.

Figure 8. Southern Avenue of Approach – Superior Observation and Fields of Fire

For MOEs, participants were provided with two scenario objectives. They were briefed that a local rebel force had recently acquired a cache of surface-to-air missiles (SAM) and participants were given the mission objective to locate and destroy the cache. The second objective would be to navigate safely to a landing zone where their avatar would be extracted via UH-60 helicopter.

Six dismounted opposing force (OPFOR) avatars wearing woodland camouflage uniforms were positioned throughout the village. OPFOR avatars were controlled by AI. They were initially standing and stationary, but would maneuver through the environment when reacting to contact with the player avatar.

The study team determined that the range at which participants engaged OPFOR soldiers would be an effective measure of how well participants were utilizing the prototype capability. To measure performance, a metric was developed where participants would receive a score that was scaled according to engagement distance for each enemy soldier killed. The study team conducted test runs of the scenario and determined that most participants would receive scores between 200–300 points per scenario based on the scaled scoring method. To maintain a 1/3 MOP, 2/3 MOE relationship, the study team assigned 500 points for accomplishing mission objectives, 300 points for destroying the cache and 200 points for navigation to the extraction point. The resulting scoring algorithm that was applied to the scenario is shown in Figure 9:

Player Score = 300 points*Cache Destroyed + 200 points*Extraction Point Reached + Σ(50*Engagement Distance/2)

Figure 9. Scenario 1 Scoring Algorithm

The scoring algorithm is specifically designed to reward players for utilizing the large effective range on the prototype rifle. Without dictating how the rifle will or

should be used, players who wish to receive high scores will know to plan an engagement that effectively uses increased observation distances and fields of fire.

3. Test Scenario 2 – Mounted Hostage Rescue

For the mounted test scenario, participants were briefed that the military was interested in fielding a replacement wheeled armored vehicle that would provide enhanced mobility and firepower compared to current Stryker variants. The enhance capabilities offered by the prototype would enable the vehicle to rapidly penetrate urban terrain while carrying a squad-sized assault force. The study team selected a pre-existing VBS3 USMC LAVA5 model to serve as the prototype vehicle for the study. This is the same vehicle that participants trained on during the training scenario. The player avatar was equipped with a standard M4 assault rifle for dismounted portions of the scenario.

The study utilized a mounted hostage rescue scenario to examine how study participants utilized the prototype vehicle and identified terrain in the VBS3 Geotypical Eastern Europe map that would support the scenario concept. The scenario brief provided to participants is included in Appendix D.

The objective compound containing the hostage was located in a rebel-controlled village consisting of three building and two trailers. The village was defended by an OPFOR consisting of twelve dismounted Taliban soldiers and three pick-up trucks with mounted machine guns. A mounted QRF equipped with anti-armor weapons was available to the QRF after fifteen minutes in order to limit the length of the scenario.

Using the proposed metric development methodology, the study team selected rescuing the hostage, safeguarding the hostage, and successfully returning to friendly controlled territory as mission objectives for the scenario.

The scenario was designed to be conducted in three phases; a mounted assault onto the objective, dismounted hostage rescue, and mounted exfiltration to friendly controlled territory.

For the first phase, participants began the scenario in a government-controlled compound approximately 5km to the southwest of the objective compound where the hostage was being held. Participants served as the vehicle driver during the initial penetration of the objective. The gunner and troop commander positions were controlled by AI. Participants had to navigate to the objective, maneuver around enemy positions and obstacles, and locate the target compound. AI forces controlled the main gun and mounted machine gun during engagements with hostile forces.

During the dismounted hostage rescue phase, participants were required to dismount the vehicle vicinity the target compound. Once dismounted, they entered the target building, eliminated any hostile forces, and rescued the hostage. Participants were then required to escort the hostage back to the LAV and prepare to egress the objective and return to friendly controlled territory.

For the final phase, the participant and hostage returned to the LAV and navigated back to friendly controlled territory. In order to encourage participants to use the speed and mobility of the prototype vehicle, the study team decided to utilize time to complete the mission as a measure of performance for the scenario. A time score was determined by starting the scenario with 600 time points and reducing the score by 1 point per second until the unit returned to friendly territory with the hostage. Participants were provided a flat rate score of 300 points for rescuing the hostage and 200 points for returning to friendly territory. 150 points were deducted for any injury to the hostage during the rescue attempt. The resulting scoring algorithm that was used to determine player score is shown in Figure 10:

Player Score = 200 points*hostage rescued – 150 points*hostage injured + 200 points*return to friendly territory + (600–mission time (seconds)) points

Figure 10. Scenario 2 Scoring Algorithm

For the scoring algorithms in both test scenario 1 and test scenario 2, certainly other scoring schemes are possible. However, it is important to note that the focus here is not on the efficacy of the particular scoring mechanism used in this study, but rather on whether or not the scoring mechanism influences behavior and increases enjoyment.

D. SURVEYS

Surveys were used to collect qualitative data to complement the quantitative data collected during gameplay. Demographic surveys were used to collect basic information about the study population. Participants completed a post-task survey that provided qualitative information related to player experiences, decision making processes, and the effects of scoring on behavior and motivation.

1. Demographic Survey

The demographic survey collected standard demographic information such as age, military service affiliation, and years of service. The survey also collected information on participants' gaming experience. The survey specifically asked for the number of hours each week that participants spent playing video games in general and military-themed games specifically. Participants were also asked to identify their favorite military-themed games for informational purposes. The demographic survey is included as Appendix A.

2. Post-task Survey

The purpose of the post-task survey was to gather qualitative data to complement game data collected by the VBS3 AAR tool. The survey employed the principle of grounding when querying participants about their game experience.

Grounding is a technique where survey respondents are asked to rate a game utilizing a game that they already play as a baseline (El-Nasr et al., 2013). In this instance, we asked participants to rate their game experience relative to their favorite military themed game.

Participants also provided a short summary of the method that they used to determine their strategies for the two scenarios and provide scaled scores on the degree to which the scoring algorithm affected their strategy, overall game experience, and willingness to participate in future studies. The post-task survey is included as Appendix B.

E. DATA COLLECTION SYSTEMS AND SOFTWARE

The VBS3 tool was used to record all scenarios. This included video play back and data on all engagements between friendly and OPFOR forces. Engagement data was then exported to Microsoft Excel in a comma separated value (CSV) format. Microsoft Excel was used to record all engagement data from VBS3 along with results from demographic and post-task surveys. Statistical analysis of engagement and survey data was conducted utilizing JMP Pro.

F. PROCEDURES

1. Prior to Experiment

Each participant was provided with a subject ID prior to arrival for the study. Once the volunteer was assigned an ID, the study ID was the only method used to identify the volunteer for the remainder of the study. The master list of volunteer and study ID combinations was secured and separated from the remainder of the study data. The volunteer was assigned to a treatment group randomly, according to his or her study ID.

Upon arriving for the study, the participant was briefed on the purpose of the study and the tasks that they would be asked to complete. Each participant was then provided with a copy of the informed consent form and provided the opportunity to read the form and ask questions prior to signing the form to indicate

his or her consent to participate in the study. After signing the consent form, the participant was asked to complete the demographic study.

The volunteer completed the VBS3 training tutorial after completing the survey. The study administrator assisted the volunteer during the study and answered any questions related to game controls to ensure that each participant had the same baseline understanding of how to operate in the game environment prior to beginning the test scenarios. Each participant completed the same training scenario and tasks to ensure uniformity.

2. During the Experiment

Each volunteer executed two test scenarios as part of the study: one unscored scenario and one scored scenario. The unscored scenario was executed first. Upon completion of training, the participant was provided a mission brief for the unscored scenario. The mission brief included information on the prototype system being researched, enemy situation, and objectives to be accomplished.

The study administrator set up two computers for the study while the participant reviewed the mission brief. The administrator utilized a workstation that was running VBS3 in administrator LVC mode. This workstation served as the server and broadcast the scenario to the participant's workstation. The scenario was then loaded onto the participant's workstation in default mode. Once the scenario was loaded, the administrator reviewed the mission brief with the study participant and answered any questions.

Once the participant stated that he or she understood the scenario and was prepared to execute, the administrator initiated the scenario and activated the AAR module on the administrator workstation. The AAR module was halted and the AAR file saved, either when the mission was successfully completed or when the participant was killed in the game.

Upon completion of the unscored scenario, the participant was provided with the mission brief for the scored scenario and was provided time to review the

mission brief, while the administrator prepared the workstations for the scored scenario. Once the scenario was loaded, the administrator reviewed the mission brief and scoring methodology with the participant and provided information on previous high scores for reference. Once the participant indicated he or she understood the scenario and were prepared to execute, the administrator initiated the scenario in the same manner as the unscored scenario. The scored scenario was recorded in the same manner as the unscored scenario.

3. After the Experiment

Upon completion of the scored scenario, participants were asked to complete the post-task survey. Once the survey was complete, the administrator provided the participants with a debriefing that included the studies design and purpose. Participants were asked not to provide information on the study to future participants.

V. RESULTS

The study utilized a combination of qualitative data from surveys and quantitative data from gameplay to investigate the impact of scoring algorithms on player behavior and experience. The study used 90% confidence as the measure of statistical significance. The study team also recorded observations of gameplay to provide further insight into how players interacted with the game environment.

A. DO CHANGES IN A SCORING ALGORITHM AFFECT PLAYER BEHAVIOR?

There is insufficient statistical evidence from this study to conclude quantitatively that the presence of scoring algorithm affected player behavior. However, qualitative data from the surveys does indicate that the scoring algorithm affected strategy. This likely has to do with a combination of the limited quantitative measures available to us for this study via VBS3 and players' lack of familiarization with the game environment.

The majority of the participants responded that the scoring algorithm had an effect on their strategy in the post-task survey. Five of the respondents stated that the scoring algorithm greatly affected their strategy (score of 9 or higher) and another eleven stated the scoring algorithm had a moderate effect (score of 5–8). One subject did not record a score for strategy effect. No significant difference appeared in the responses between subjects who were scored on the dismounted scenario (3 high, 6 moderate) and the mounted scenario (2 high, 5 moderate). A t-test of the survey data determined with a 10% confidence interval that the mean strategy effect of a scoring algorithm would be between 5.42 and 7.78 on an 11-point scale. The upper and lower bounds of the confidence interval both fall within the moderate effect category as depicted in Table 3 and Figure 10.

Table 3. T-Estimate of Mean Effect of Scoring Algorithm on Strategy – 90% Confidence Interval

t-Estimate: Mean	
All	
Mean	6.60
Standard Deviation	3.05
LCL	5.42
UCL	7.78
Scenario 1 Scored	
Mean	7.40
Standard Deviation	2.50
LCL	5.95
UCL	8.85
Scenario 2 Scored	
Mean	6.44
Standard Deviation	2.96
LCL	4.61
UCL	8.28

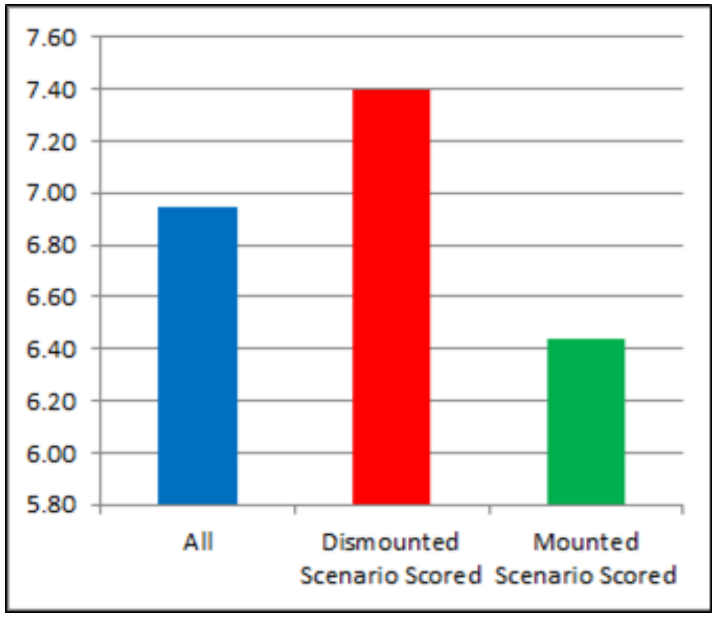


Figure 11. Scoring Effect on Strategy

It is difficult to verify this strategy effect when comparing actual gameplay from the two scenarios. When asked to describe how they developed their strategy on the scored scenario, eight of the participants provided responses consistent with achieving a high score (3 from the dismounted scenario, 5 from the mounted scenario). This is only slightly higher than the six (3 from the dismounted scenario, 3 from the mounted scenario) who described strategies that would produce a high score. This suggests that the capabilities provided by the prototype systems also had an effect on strategy formulation.

Data collected from gameplay is insufficient to determine that the scoring algorithms significantly affected gameplay for the study as a whole or for either scenario. However, as stated earlier, this would require a quantitative measure of behavior (or strategy) that is beyond the capability of VBS3. We provide more detail on what measures were available in the following section.

1. Scenario 1 – Dismounted Raid

Four measures compared gameplay between participants who were scored on their execution of the dismounted raid scenario and those who were not scored: mission success, mission score, average engagement distance, and first engagement distance. Participants who were scored when executing the dismounted scenario generally had greater average engagement distances and first engagement distances. However, participants who were not scored achieved a higher average score and successfully complete the mission more often. The difference in performance was not enough to be statistically significant for any measurement.

a. Mission Success

Only four participants successfully completed the dismounted scenario. The study team defined mission success as locating and destroying the SAM cache and arriving at the extraction point. Only one of the successful participants had a scoring algorithm applied. Three of the successful participants were unscored. Of the four players who successfully completed the scenario, three were the

volunteers who identified themselves as playing more than ten hours of video games a week. The other successful player had a high level of proficiency in the VBS3 game environment.

b. Scenario Score

Participants who were not scored when executing the dismounted had a higher average mean scenario score (269.89) when compared to those who were provided a scoring algorithm prior to execution (174.82). This is partially due to the fact that three of the four players who successfully completed the scenario were in the unscored group. However, due to the high degree of variability in the scores, the data is insufficient to state that there is a statistically significant difference in the mean scores between players who are scored and players who are not. Table 4 shows the 90% confidence interval for mean scenario scores for all iterations of the scenario and for the scenario under scored and unscored conditions. Table 5 shows a paired t-test of the mean scenario scores for when the scenario is either scored or not scored, which shows that the difference between the two means is not significantly different. Figure 11 is a graphical depiction of mean scores for the dismounted scenario.

Table 4. T-Estimate of Mean Scores for Scenario 1 – 90% Confidence Interval

t-Estimate: Mean	
ALL	
Mean	224.04
Standard Deviation	251.09
LCL	126.95
UCL	321.12
Scenario Scored	
Mean	174.82
Standard Deviation	240.24
LCL	35.55
UCL	314.08
Scenario Unscored	
Mean	269.89
Standard Deviation	271.70
LCL	112.39
UCL	427.38

Table 5. T-Test: Paired Two Sample for Means – 90% Confidence

	Scored	***Unscored***
Mean	174.82	269.89
Variance	57713.51	73821.31
Observations	10	10
df	9	
t Stat	-0.79	
P(T<=t) two-tail	0.45	
t Critical two-tail	1.83	

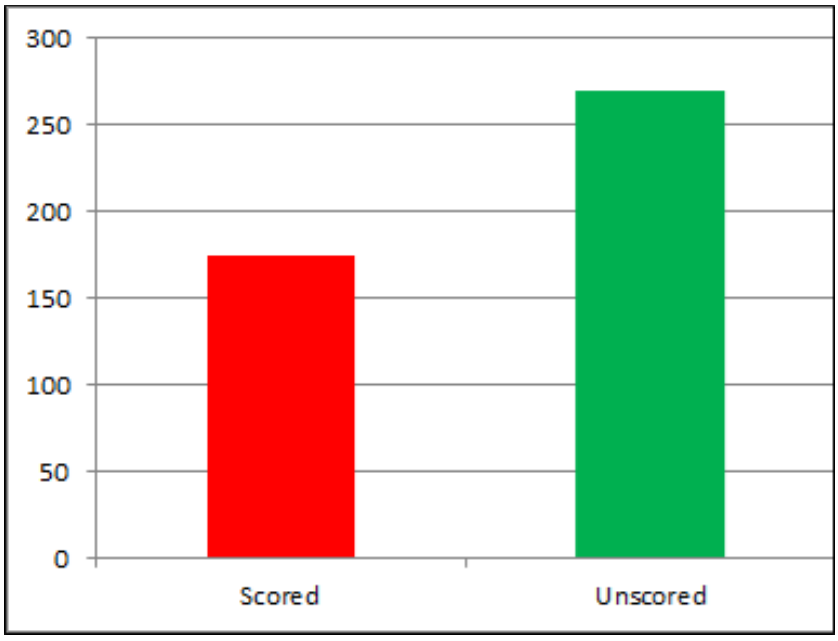


Figure 12. Mean Scores for Scenario 1

c. Mean Engagement Distance

Players executing the dismounted scenario with a scoring algorithm had a higher mean engagement distance (108.22 meters) compared to those who did not have the scoring algorithm applied (73.32 meters) as depicted in Figure 13. As with scoring, the data reflects a high degree of variability. A paired t-test comparison of players who execute the scenario with and without the scoring algorithm applied yields a t-stat of 0.97 and a p-value of 0.18 (Table 7). This is not sufficient statistical evidence to conclude that players have a higher average engagement distance when a scoring algorithm is applied. Ninety percent confidence intervals for the mean engagement distance of the scenario are shown in Table 6.

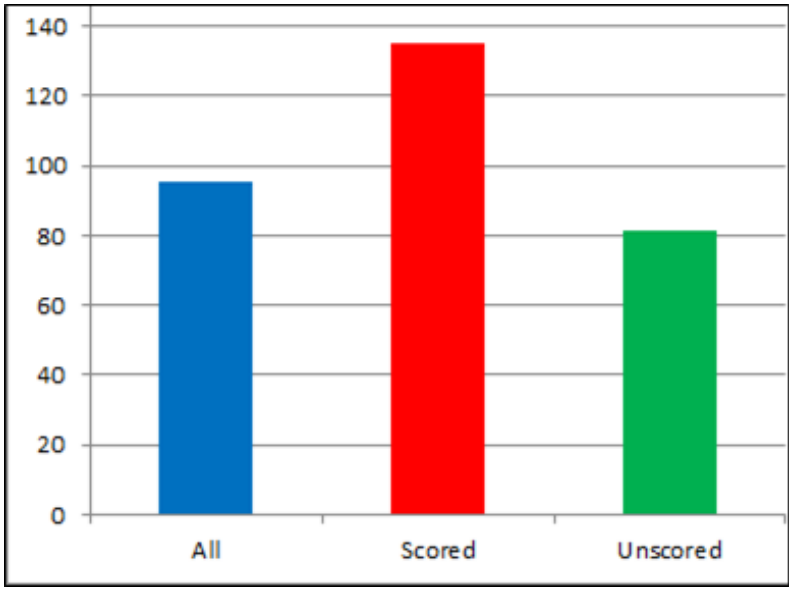


Figure 13. Mean Engagement Distance (meters)

Table 6. T-Estimate: Mean Engagement Distance – 90% Confidence

All	
Mean	90.77
Standard Deviation	75.48
LCL	61.58
UCL	119.95
Scored	
Mean	108.22
Standard Deviation	100.35
LCL	50.04
UCL	166.39
Unscored	
Mean	73.32
Standard Deviation	35.80
LCL	52.57
UCL	94.07

Table 7. T-Test: Paired Two Sample for Mean Engagement Distance – 90% Confidence

	Scored	*Unscored*
Mean	108.22	73.32
Variance	10070.42	1281.39
Observations	10	10
Hypothesized Mean	0	
df	9	
t Stat	0.97	
P(T<=t) one-tail	0.18	
t Critical one-tail	1.38	

If data from the three players who did not kill any OPFOR soldiers are excluded, the difference between the mean engagement distances of players who are scored and unscored is greater. However, there is still not enough statistical evidence to conclude that the two means are different as shown by the overlap of 90% confidence intervals in Table 8.

Table 8. T-Estimate: Mean Engagement Distance (No Kills Excluded) – 90% Confidence

All	
Mean	95.54
Standard Deviation	74.38
LCL	65.95
UCL	125.13
Scored	
Mean	135.27
Standard Deviation	93.62
LCL	72.56
UCL	197.98
Unscored	
Mean	81.46
Standard Deviation	26.36
LCL	65.12
UCL	97.81

d. First Kill Engagement Distance

The intent for the dismounted scenario is for players to utilize the increased range of the prototype rifle to attrite the enemy force before assaulting the village. In most instances, close range engagements will occur inside the village. These engagements will have a large degree of influence by decreasing the overall mean engagement distance. Since engagement distances inside the village will be similar for both scored and unscored players, these engagements will decrease the difference between average engagement distances between the two groups. Therefore, it is also useful to look at only the distance of the first engagement with OPFOR when evaluating how well a player utilizes the additional range provided by the prototype system. When comparing only the distance of the engagement for the first OPFOR soldier killed, players who had the scoring algorithm applied had a mean engagement distance of 124.25 meters compared to 83.16 meters for those who did not (Figure 14).

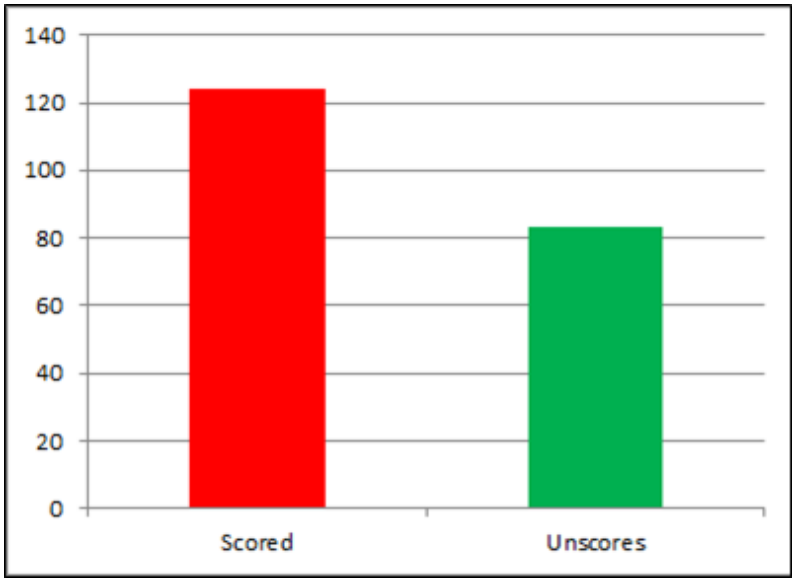


Figure 14. Mean First Engagement (meters)

The data on first engagements is still not statistically significant. However, if the confidence level were decreased to 83%, the results would become significant.

Ninety percent confidence intervals for the mean distance of first engagements are shown in Table 9. A paired t-test comparing the mean first engagement distance when the scoring mechanism is applied with the mean distance with no scoring mechanism is present is in Table 10.

Table 9. T-Estimate: Mean First Engagement Distance – 90% Confidence

All	
Mean	103.71
Standard Deviation	86.43
LCL	70.29
UCL	137.12
Scored	
Mean	124.25
Standard Deviation	109.53
LCL	60.76
UCL	187.74
Unscored	
Mean	83.16
Standard Deviation	53.26
LCL	52.28
UCL	114.04

Table 10. T-Test: Paired Two Sample for Mean First Engagement Distance – 90% Confidence

	Scored	*Unscored*
Mean	124.25	83.16
Variance	11997.01	2837.06
Observations	10	10
Hypothesized Mean Difference	0	
df	9	
t Stat	1.021089	
P(T<=t) one-tail	0.166937	
t Critical one-tail	1.383029	

When results from players who did not kill any OPFOR are excluded, the mean first engagement distances of players with the scoring algorithm applied and those without become 135.27 meters and 92.40 meters respectfully as shown in Figure 15 and the t-test confidence interval in Table 11.

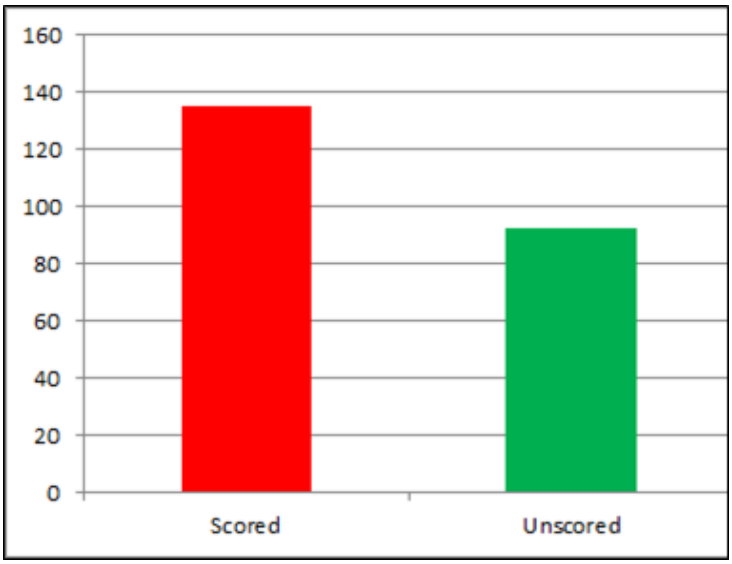


Figure 15. Mean First Engagement Distance – No Kills Excluded (meters)

Table 11. T-Estimate: Mean First Engagement Distance (No Kills Excluded) – 90% Confidence

All	
Mean	122.01
Standard Deviation	80.62
LCL	87.87
UCL	156.14
Scored	
Mean	135.27
Standard Deviation	93.62
LCL	72.56
UCL	197.98
Unscored	
Mean	92.40
Standard Deviation	47.24
LCL	63.12
UCL	121.68

2. Scenario 2 – Mounted Hostage Rescue

Two measures compared gameplay between participants who were scored on their execution of the mounted hostage rescue scenario and those who were not scored: mission success and scenario score. Participants who were not scored had more mission successes and a higher mean score, although the difference in means was not statistically significant.

a. Mission Success

Nine players successfully completed the mounted hostage rescue scenario, including three of the four players who successfully completed the dismounted raid scenario. Four of the successful players had the scoring algorithm applied, while five did not. Two of three players who were identified as gamers were successful. Seven of seventeen non-gamers successfully completed the scenario. No apparent difference emerged between mission success of the scored group vs the unscored group.

b. Scenario Score

Players who had a scoring algorithm applied to the mounted hostage rescue scenario had a lower mean score when compared to players who were not scored. Players who were scored had a mean score of 317.70 compared to an average score of 411.70 for players who were not scored (Figure 16). 90% confidence intervals are depicted in Table 12 depicts 90% confidence intervals for player scores. There was a high degree of variability in the data. A major cause of this is that players that did not successfully complete the scenario generally did not receive any points. Due to the high degree of variability in the data and the small sample size, the differences in mean scores was not sufficiently significant to state there is a difference in performance between the two groups as depicted in the paired t-test in Table 13.

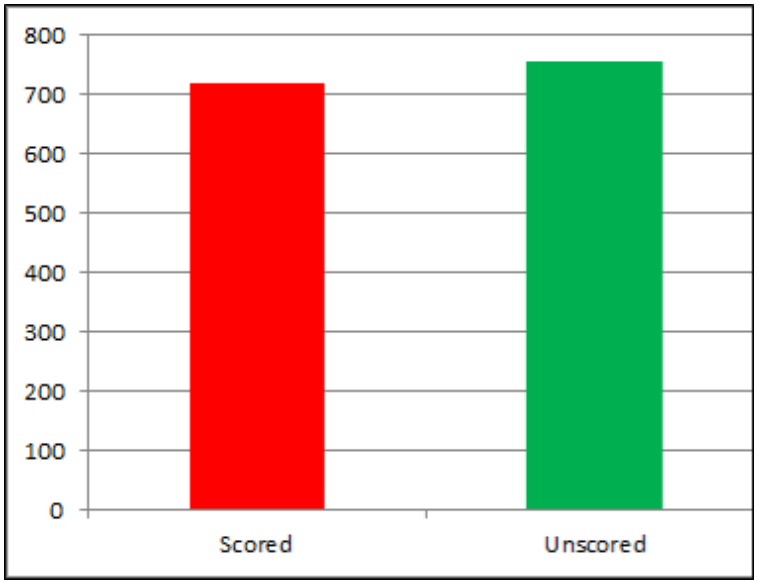


Figure 16. Mean Scores – Mounted Scenario

Table 12. T-Estimate: Mean Score – 90% Confidence

All	
Mean	364.70
Standard Deviation	359.91
LCL	225.54
UCL	503.86
Scored	
Mean	317.70
Standard Deviation	361.11
LCL	108.37
UCL	527.03
Unscored	
Mean	411.70
Standard Deviation	371.68
LCL	196.24
UCL	627.16

Table 13. T-Estimate: Mean Score, Successes Only – 90% Confidence

	Scored	*Unscored*
Mean	317.70	411.70
Variance	130402.23	138147.34
Observations	10	10
Hypothesized Mean Difference	0	
df	9	
t Stat	-0.53	
P(T<=t) two-tail	0.61	
t Critical two-tail	1.83	

One of the reasons the mean score for the unscored group was greater than the mean score of the scored group is that the unscored group had more successes. If you discount the failed missions and consider the mean score of only successful missions, little difference separated the scored and unscored players' performances. Evaluating only successful missions, the unscored group had a

mean score 754.40 compared to 719.25 for the scored group. This difference in means is statistically insignificant, which is shown by the overlap in 90% confidence intervals in Table 14.

Table 14. T-Estimate: Mean Score, Successes Only – 90% Confidence

All	
Mean	738.78
Standard Deviation	85.58
LCL	685.73
UCL	791.82
Scored	
Mean	719.25
Standard Deviation	88.83
LCL	614.73
UCL	823.77
Unscored	
Mean	754.40
Standard Deviation	89.68
LCL	668.90
UCL	839.90

B. DO CHANGES IN A SCORING ALGORITHM AFFECT PLAYER ENJOYMENT?

The assessment of how scoring algorithms affect a player's enjoyment of the game environment was conducted utilizing qualitative data only. The study design did not support gathering sufficient data to conduct a quantitative analysis of player behavior that could be associated with enjoyment. User input on the post-task survey was used for this analysis. Specific questions examined related to users' impressions of the overall game experience as well as feedback on how the scoring mechanism affected their enjoyment and willingness to participate in future ESP studies.

1. Overall Game Experience

The survey asked users to rate their overall game experience compared to their military themed game. The question utilized grounding to establish users' favorite games as a baseline to provide context to the response. The study team chose to utilize this grounding technique because ESP will compete for playing time with commercial games that are created for enjoyment purposes. In retrospect, this may not have been a useful strategy for this study, as a large proportion of the study population did not play video games in their free time.

Given that the scenarios for this study were created in an environment, it is reasonable to expect that users would not rate the game experience as superior to their favorite commercial game and may likely rate their experience as being inferior to that of their favorite game. However, many players did express that they appreciated many features of the VBS3 game environment to include the fatigue modeling and realistic physics. Based on this the study expected to find that participants rated their game experience as roughly equal to that of their favorite games. The survey responses did indicate this was the case. Participants rated their game experience with a mean score of 6.40 compared to a rating of 6 that would be an equal experience to their favorite game. This difference was statistically insignificant from a population mean score of 6 as depicted in the t-test in Table 15.

Table 15. T-Test: Mean Game Experience Rating – 90% Confidence

All	
Mean	6.30
Standard Deviation	2.54
Hypothesized Mean	6
df	19
t Stat	0.53
P(T<=t) two-tail	0.60
t Critical two-tail	1.73

Players who were scored on the dismounted scenario rated their game experience slightly higher (6.50) than players who were scored on the mounted scenario (6.10). Neither of these ratings differed significantly from the hypothesized mean of 6 as shown in Table 16.

Table 16. T-Test: Mean Game Experience Rating by Test Group – 90% Confidence

Dismounted Scenario Scored	
Mean	6.50
Standard Deviation	2.80
Hypothesized Mean	6
df	9
t Stat	0.56
P(T<=t) two-tail	0.59
t Critical two-tail	1.83
Mounted Scenario Scored	
Mean	6.10
Standard Deviation	2.38
Hypothesized Mean	6
df	9
t Stat	0.13
P(T<=t) two-tail	0.90
t Critical two-tail	1.83

2. Scoring Mechanism Effect on Enjoyment

Players rated the degree to which scoring contributed to their enjoyment of the game on an 11-point scale with 1 meaning that scoring greatly decreased their enjoyment, 6 meaning it had no effect, and 11 meaning that scoring greatly enhanced their level of enjoyment.

Player responses indicated that the presence of a scoring mechanism did contribute to enjoyment. Survey responses had a mean score of 7.25 with a

standard deviation of 2.22. This was sufficient to state that scoring contributed to enjoyment with a 90% confidence level. A t-test, 90% confidence interval indicates that the true mean for level of enjoyment is between 6.39 and 8.11 (Table 17). This would be equivalent to a slight to moderate contribution to enjoyment level.

Table 17. T-Estimate: Mean Contribution to Enjoyment Confidence Interval – 90% Confidence

All	
Mean	7.25
Standard Deviation	2.22
LCL	6.39
UCL	8.11

Players who were scored on the dismounted scenario generally responded that scoring had a greater effect on scoring compared to players who were scored on the hostage rescue scenario. Players who were scored on the dismounted scenario had a 7.80 mean response while players scored on the hostage rescue scenario had a mean response of 6.70. The responses are not significantly significant when compared to each other. However, when the responses of the players scored on the hostage rescue scenario are compared to the hypothesized mean of 6.0, there is not enough statistical evidence to conclude that the scoring mechanism on that scenario contributed to player enjoyment. The t-tests for the mean level that the scenario scores provided to enjoyment for each scenario are shown in Table 18.

Table 18. T-Test: Mean Contribution to Enjoyment, scenario Scores – 90% Confidence

Dismounted Scenario Scored	
Mean	7.80
Standard Deviation	1.55
Hypothesized Mean	6.00
df	9.00
t Stat	3.67
P(T<=t) two-tail	0.01
t Critical two-tail	1.83
Mounted Scenario Scored	
Mean	6.70
Standard Deviation	2.71
Hypothesized Mean	6.00
df	9.00
t Stat	0.82
P(T<=t) two-tail	0.44
t Critical two-tail	1.83

That the mean enjoyment rating for one scenario is significant while the other is not could indicate that not all scoring mechanisms contribute to enjoyment equally. Notably, one player rated the scoring mechanism on the mounted scenario as a 1 meaning it greatly detracted from enjoyment. This response is an outlier and the only response that did not indicate at least a neutral effect on enjoyment. If this response is discounted, the mean enjoyment rating for players scored on the mounted scenario increases to 8.89, and is statistically greater than the hypothesized mean of 6.0 (Table 19).

Table 19. T-Test: Mean Contribution to Enjoyment, Scenario 2 Scored With Outlier Excluded – 90% Confidence

*S*cenario 2 Scored - Outlier Excluded	
Mean	8.89
Standard Deviation	2.32
Hypothesized Mean	6
df	8
t Stat	3.74
P(T<=t) two-tail	0.01
t Critical two-tail	1.86

3. Scoring Effect on Willingness to Participate in Future Studies

Participants rated the likely affect the presence of a scoring mechanism and leader boards would have on their willingness to participate in future ESP style studies. Similar to the question on scoring's effect on enjoyment, players provided their response on an 11-point scale with 1 meaning they would be much less likely to participate, 6 indicating no effect, and 11 meaning that they would be much more likely to participate.

Player responses indicated that the presence of a scoring mechanism would make them more likely to participate in ESP studies. The mean response was 7.90 with a standard deviation of 2.77. This is sufficient statistical evidence to indicate a positive effect on likelihood that players will participate in future studies as illustrated by the t-test in Table 20.

Table 20. T-Test: Mean Rating Scoring Effect on Willingness to Participate in Future Studies – 90% Confidence

All	
Mean	7.90
Standard Deviation	2.77
Hypothesized Mean	6
df	19
t Stat	3.07
P(T<=t) two-tail	0.01
t Critical two-tail	1.73

Players who were scored on the mounted scenario responded with a mean score of 7.70 that scoring mechanisms would make them more likely to participate in studies, while players scored on the hostage rescue scenario provided a mean response of 8.10. These responses are not statistically different from each other and both provide sufficient statistical evidence to conclude that respondents from both groups would be more likely to participate as shown by Table 21.

Table 21. T-Test: Mean Rating Scoring Effect on Willingness to Participate in Future Studies, Scenario Scores – 90% Confidence

Dismounted Scenario Scored	
Mean	7.70
Standard Deviation	2.26
Hypothesized Mean	6
df	9
t Stat	2.38
P(T<=t) two-tail	0.04
t Critical two-tail	1.83
Mounted Scenario Scored	
Mean	8.10
Standard Deviation	3.31
Hypothesized Mean	6
df	9
t Stat	2.00
P(T<=t) two-tail	0.08
t Critical two-tail	1.83

VI. DISCUSSION AND CONCLUSIONS

This study proposed a methodology to generate scoring algorithms for use in ESP based research studies supporting future acquisition programs. The study provided qualitative evidence from user surveys that the presence of a scoring mechanism affected their strategy when executing scenarios in an environment similar to future ESP game environments. The study also provided qualitative evidence that the presence of a scoring algorithm and leader boards would contribute to player enjoyment and likelihood of participating in future studies. The study did not provide sufficient quantitative evidence from gameplay to support the qualitative data from the player surveys. This is partly due to limitations of the study that can be improved upon by future research efforts.

1. Study Limitations

This study had a number of limitations that could be improved upon in future studies. One issue is that the study did attract the intended target population of volunteers who frequently play video games in their free time. This means that the study population may not be representative of the population who will be expected to participate in future ESP studies.

Another limitation is that the study was not available online. This limited the volunteers to one opportunity to play each scenario. Allowing participants to play scenarios multiple times would have enabled the study team to observe how player behavior evolved over time. This observation would have identified if there were a difference in how player behavior evolved when a scoring algorithm was present. This would have also allowed the collection of quantitative data related to player enjoyment. This could have been done by observing if players who were presented with scoring algorithms were more or less likely to play the scenarios multiple times and if there was a difference in the number of times they played the scenarios compared to players who were not provided with scoring algorithms.

The study team lacked the capability to obtain source-level game metrics from VBS3. This data would have allowed the study to provide real-time scoring updates to players during the scenario. This would have also supported development and testing of complex scoring algorithms more tightly coupled to the design question with more measures of performance to evaluate how players were utilizing the prototype systems.

2. Future Work and Recommendations

A study should be developed where a game environment is available online and volunteers are able to participate multiple times. The study should incorporate the proposed scoring methodology and collect quantitative data on how player behavior changes over time when the scoring mechanism is applied.

Second, the study should be repeated with a pool of volunteers that more closely represents the population who will participate in ESP studies. This will provide more accurate insights into how scoring mechanisms will affect the target audience.

The methodology should be altered to determine if altering the relative importance of game objectives and measures of performance in the scoring algorithm affects game play and user enjoyment. Multiple test groups should be included, with each group being presented with a different scoring algorithm. This will help determine the best methodology for developing scoring algorithms for future studies. The study should also utilize a range of possible prototypes and mission types to study if the scoring algorithm can be used to influence desirable player behaviors for multiple scenarios.

APPENDIX A. DEMOGRAPHIC SURVEY

VBS3 Scenario Task
Demographic Survey

Subject #:_____ Date:_____

1. Age:_____
2. Gender: Male Female
3. Are you currently serving in the Armed Forces: Yes No
    a. Branch:_____
    b. Years of Service: _____
    c. Highest Rank:_____
    d. Have you deployed to a combat zone (receipt of Imminent Danger Pay)?
        No (skip to e.) Yes (i-iii below)
        i. Number of deployments / total months deployed _____
        ii. Date of return from last deployment_____
4. How many hours per week do you play video games?_____
5. How many hours per week do you play military themed video games?_____
6. What are your top 3 favorite military themed games?_____

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. POST-TASK SURVEY

VBS3 Scenario Task
Post Task #1 Survey

Subject #:_____ Date:_____

7. Compared to your favorite military themed game, how would you rate you game experience?
   Scale(1-11, 1 being "much worse" 6 being "about the same" and 11 being "much better") _____

8. How would you rate ease of using the game controls? (1-11, 1 being "very difficult" and 11 being "very easy")_____

9. How did you determine your strategy when completing the first scenario?_____

   _____

10. How did you determine your strategy when completing the second scenario?_____

    _____

11. How much did the scoring mechanism affect your strategy during the scenario?
    Scale(1-11, 1 being "no effect" and 11 being "great effect") _____

12. How much did the scoring mechanism affect your enjoyment of the game?
    Scale(1-11, 1 being "much less enjoyment," 6 being "no effect" and 11 being "much greater enjoyment") _____

13. Do you feel that providing scores / leader boards would make you more or less likely to participate in a ESP study?

    Scale(1-11, 1 being "much less likely," 6 being "no effect" and 11 being "much more likely") ____

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. SCENARIO 1 MISSION BRIEF – DISMOUNTED RAID

For this scenario, you have been equipped with a prototype M4 variant that with a maximum effective range of 500 meters (43% increase over current M4). The Army is interested in receiving soldier feedback on the effect of providing units with a primary weapon system that provides increased range with comparable performance to current weapon systems in close quarters combat.

You have deployed to the country of Sarhani as an advisor to the Sarhani military in their campaign against rebel forces. Intel reports that rebels have stolen a cache of anti-aircraft weapons and are preparing to use them against coalition forces utilizing an air corridor between Iguana and Paraiso. You have received a mission to raid the rebel stronghold and destroy the cache before the weapons can be employed. Intelligence believes the cache is in a walled compound on the south side of the main east-west running road.

You have infiltrated rebel held territory from the village of Dolores and are currently located at a release point east of the rebel stronghold. There are two roads into the village. An east-west road that is located directly to your south that runs through the city. There is also a north-south running road that leads north out of the stronghold to the rebels' self-proclaimed capital and the seat of their militia.

The terrain to your south consists of high ground with gently rolling terrain sloping generally down into the rebel village. There is sporadic cover and concealment and observation and fields of fire are generally clear and are limited mainly by micro terrain.

The terrain to your north consists of a shallow valley and a grove of hardwood trees. There is ample cover and concealment from vegetation and terrain to the eastern edge of the village. Observation and fields of fire are generally restricted due to vegetation.

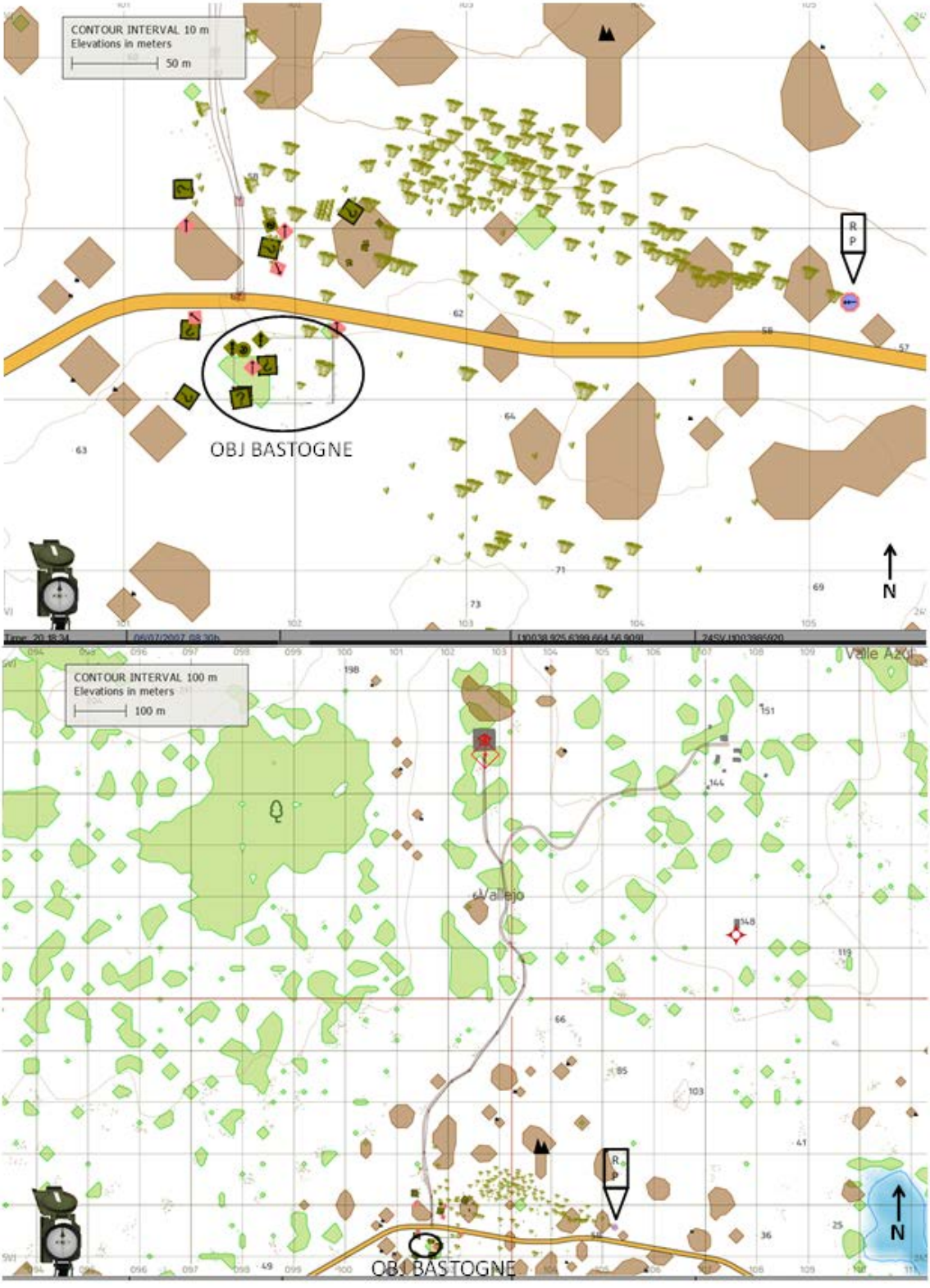
Reports indicate that the stronghold is held by a squad size element of rebel forces that are prepared to defend in place. Mounted forces with heavy machine guns are preparing to reinforce from Vallejo and are expected to arrive within the next 15–20 minutes. There are no reports of civilians in the area.

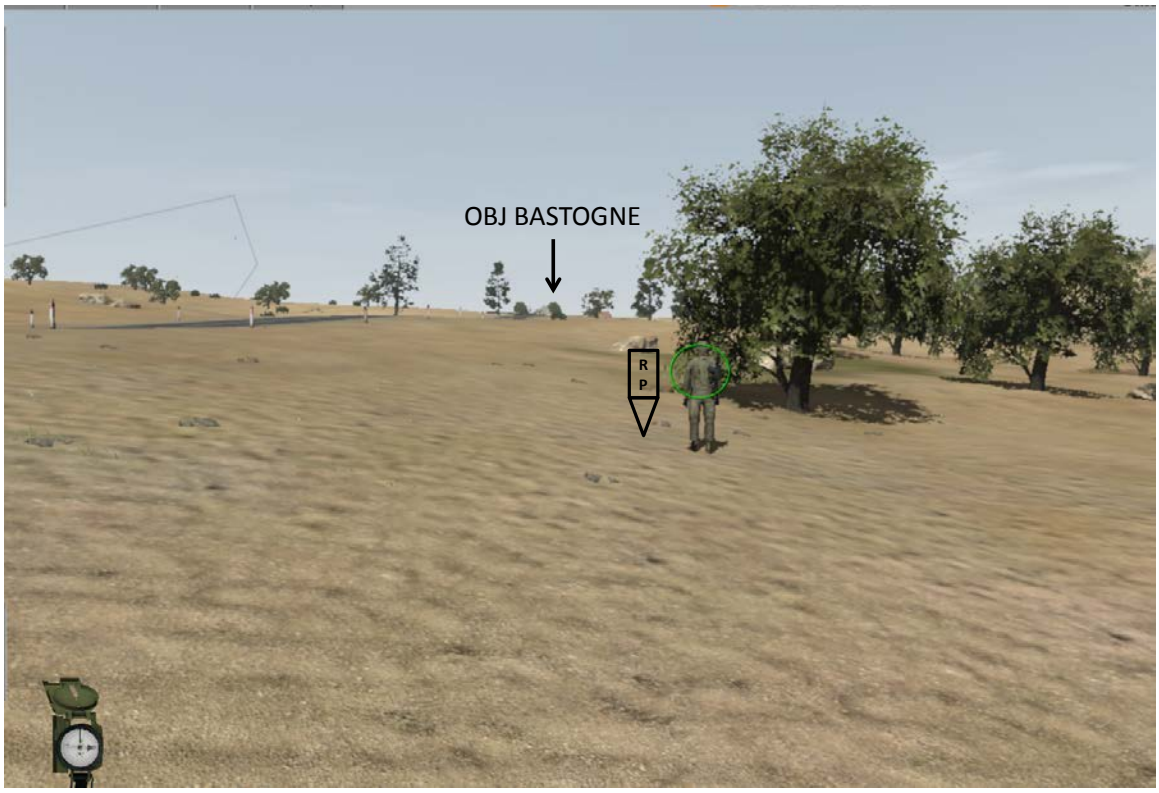
Your mission is to locate and destroy the weapons cache and then move to an extraction point at 24S VJ 10018 85620 where you will be extracted via UH-60.

You have been equipped with a suppressed M4 prototype with maximum effective range of 500meters and 8x 30-round magazines, a suppressed M9 with 3x 15-round magazines, 2x Satchel charges, 2x Fragmentation grenades, and 2x Smoke grenades.

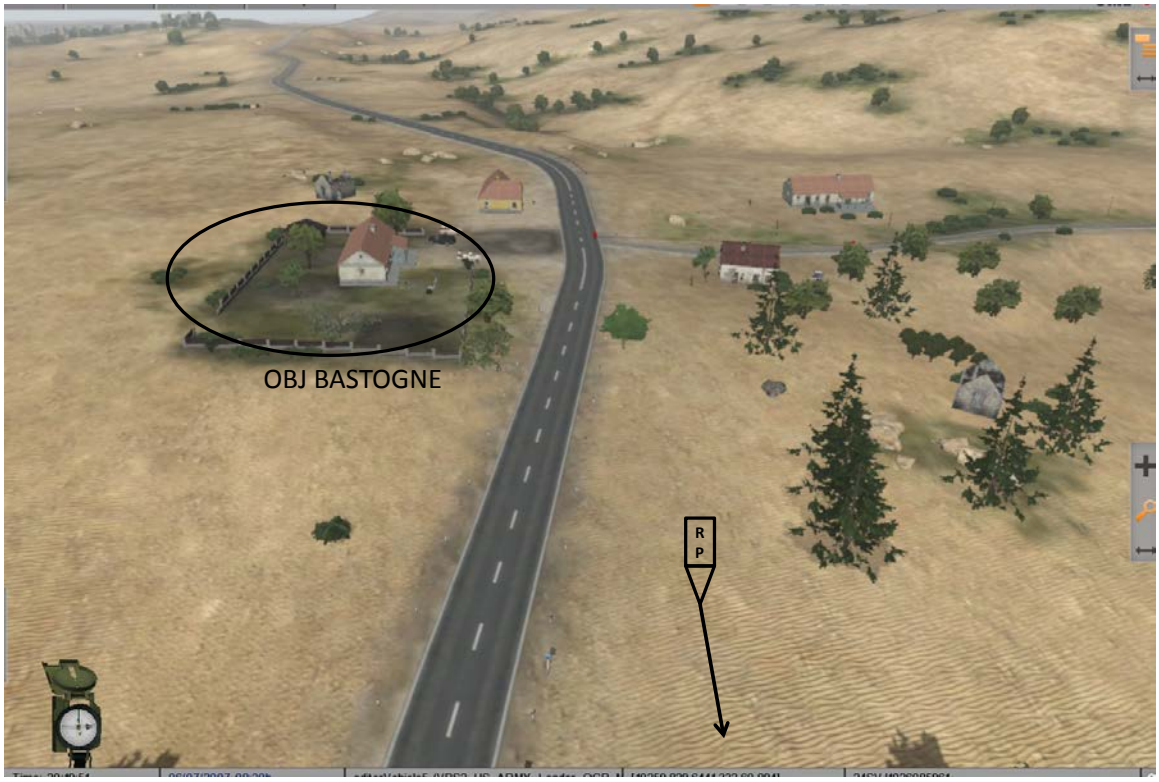
This scenario will be scored to evaluate the performance of players equipped with the prototype system. You will be awarded a flat rate score for destroying the enemy cache and successfully navigating to the exfil point. The score for killing rebel soldiers will be scaled according to engagement distance. The score will be calculated as follows.

Destroy Cache:	300 points
Arrive extraction point:	200 points
Enemy killed:	(50 * engagement distance / 100) points


CONTOUR INTERVAL 10 m
Elevations in meters
50 m
OBJ BASTOGNE
N
CONTOUR INTERVAL 100 m
Elevations in meters
100 m
Valle Azul
Vallejo
OBJ BASTOGNE
N



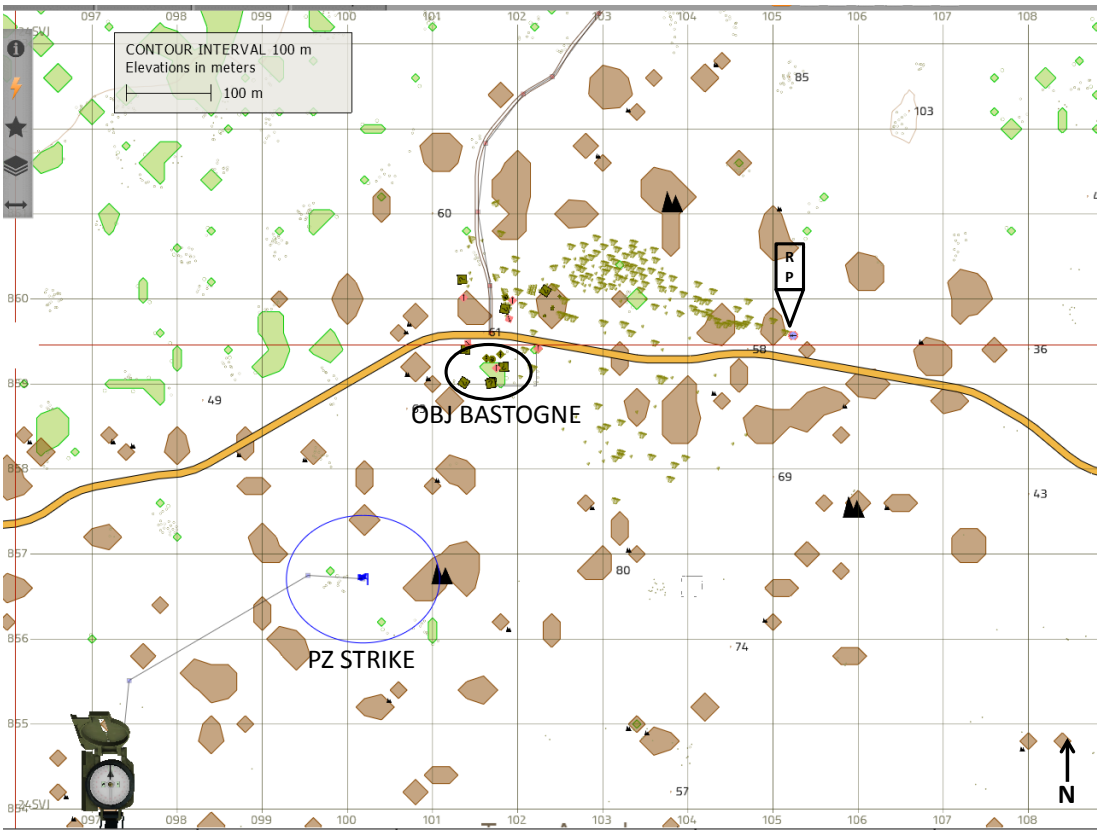
N →

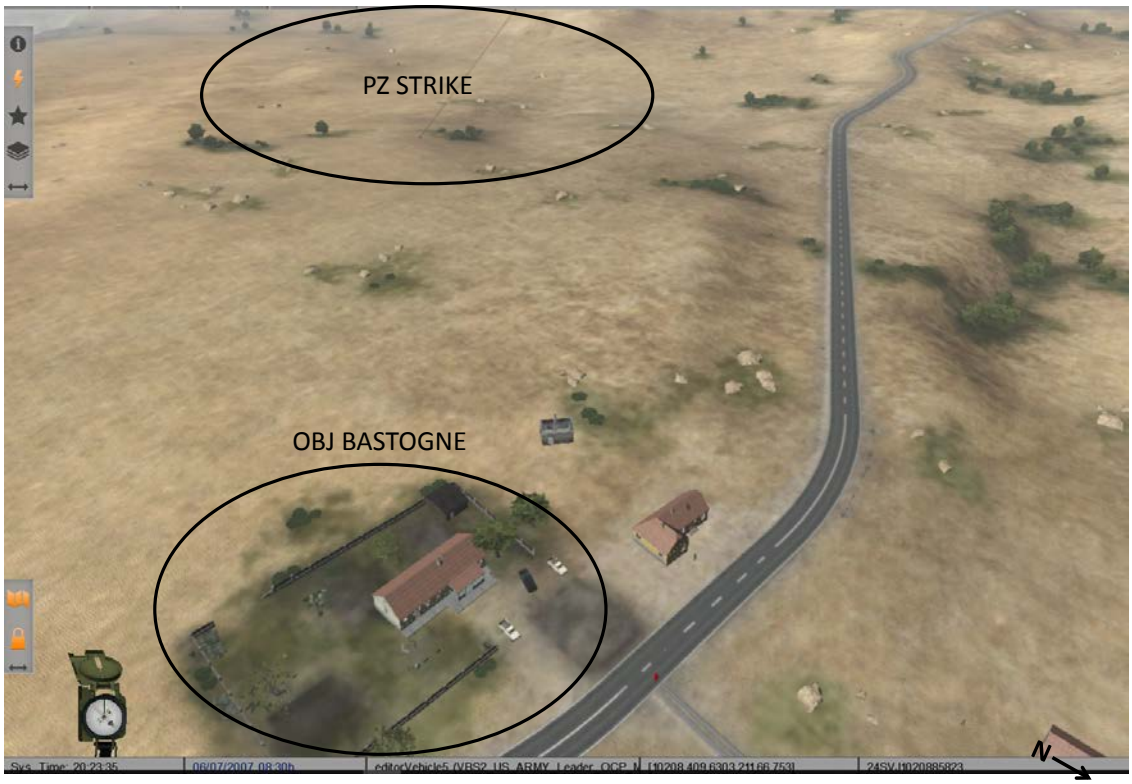


N →


R
P
OBJ BASTOGNE

N


CONTOUR INTERVAL 100 m
Elevations in meters
100 m
R
P
OBJ BASTOGNE
PZ STRIKE
N


PZ STRIKE
OBJ BASTOGNE
N

APPENDIX D. SCENARIO 2 BRIEF – MOUNTED HOSTAGE RESCUE

In this scenario, the military is interested in fielding a wheeled, armored troop transport that could serve as an armored assault platform in urban combat. The proposed vehicle would provide enhanced firepower and mobility compared to current Stryker variants while providing armor protection that is at least equal to that provided by the Stryker. It would be able to penetrate into hostile urban terrain while carrying a 8-man assault force.

Your team is deployed to Gorgas training and advising the Gorgan Army to defeat an uprising of an armed militia backed by the Atropian government. Militia forces have captured an aid worker and are holding him for ransom. Intelligence reports that the hostage is being held in a rebel controlled village to the North of the town of Oak Grove. The hostage is believed to be located in a walled compound in the center of the village. Rebel forces are preparing to move the hostage via water to Atropia within the next 12 hours.

Your team is located in a government compound near the Oak Grove airfield. Your mission is to conduct a raid on the rebel village and rescue the aid worker before the rebel forces can transport him to Atropian soil. Your team is equipped with wheeled armored transportation with a 25mm cannon. Your armor can provide protection against direct fire up to .50-caliber ammunition.

Intelligence reports that the village is defended by a reinforced squad with technical vehicles, assault rifles, and medium machine guns. There are prepared defenses along high speed avenues of approach into the village. Reporting indicates that a rebel platoon size element equipped with technical vehicles with mounted anti-armor weapons is preparing to reinforce from Trenton, a rebel held village Northeast of Oak Grove. Reinforcements are expected to arrive within the next 15–20 minutes.

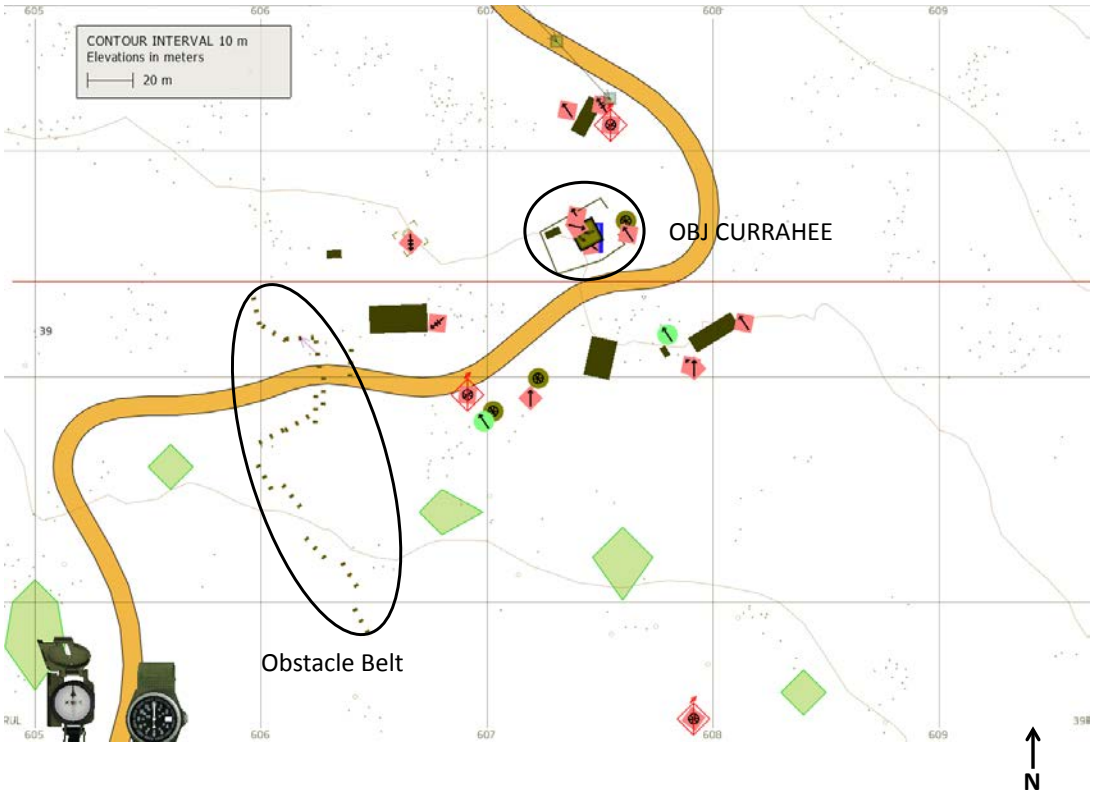
Your team has been equipped with a prototype wheeled, armored transport with 25mm main gun and 7.62 coaxial machine gun on a 360-degree rotating turret. You are equipped with a M4 carbine with PEQ-15 sights and 8x 30-round magazines, a M9 pistol with 3x 15 round magazines, 2x fragmentation grenades, and 2x smoke grenades.

This scenario will be scored to evaluate the performance of players equipped with the prototype system. You will be awarded a flat score for rescuing the hostage without injury and for maneuvering back to friendly territory. There will be a penalty assessed if the hostage is wounded during the rescue attempt. The score will also be adjusted according to the total time required to complete the mission. The score will be calculated as follows.

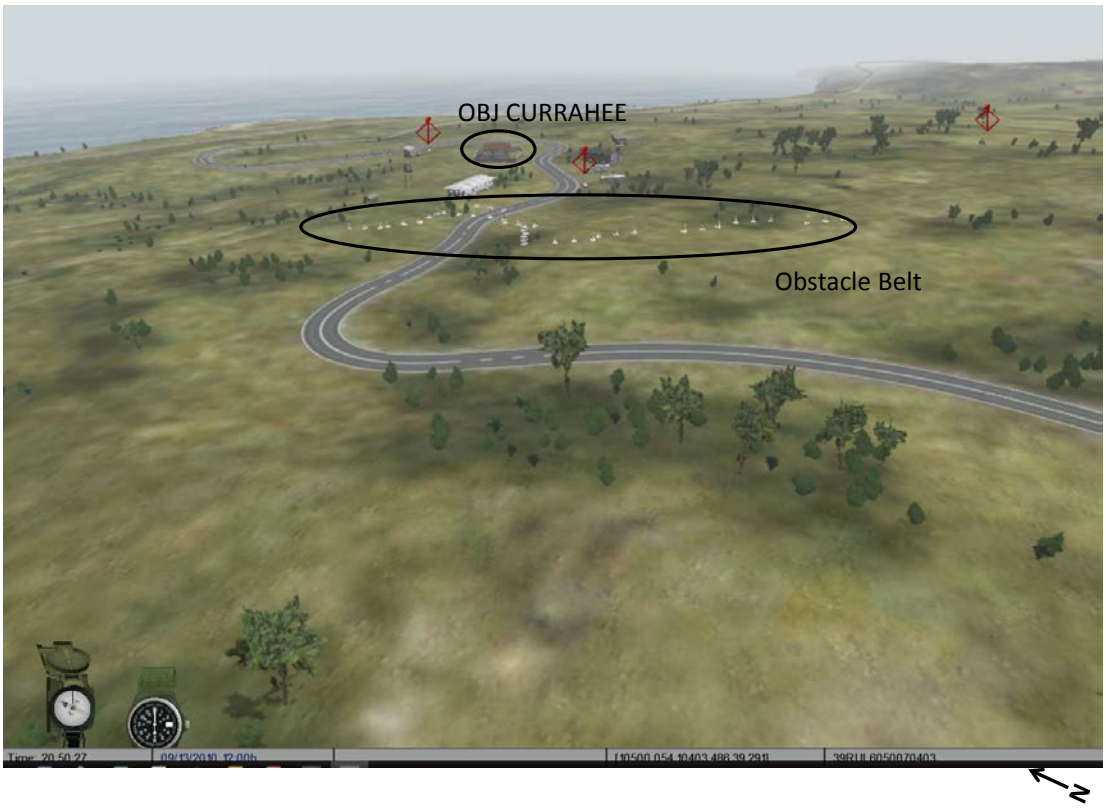
Rescue hostage:	300 points
Re-enter friendly territory:	200 points
Injury to hostage:	(-)150 points
Time factor (mission time in seconds):	600-mission time points

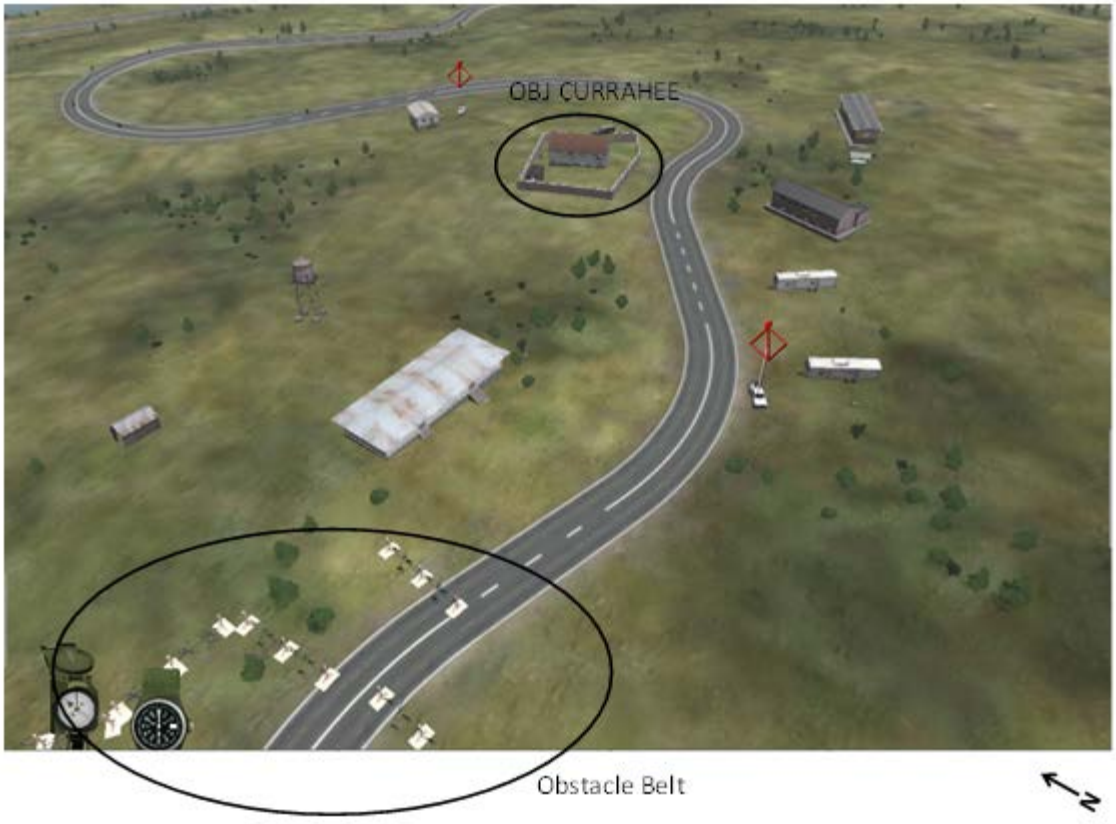

CONTOUR INTERVAL 100 m
Elevations in meters
100 m
OBJ CURRAHEE
Oak Grove Airfield
N

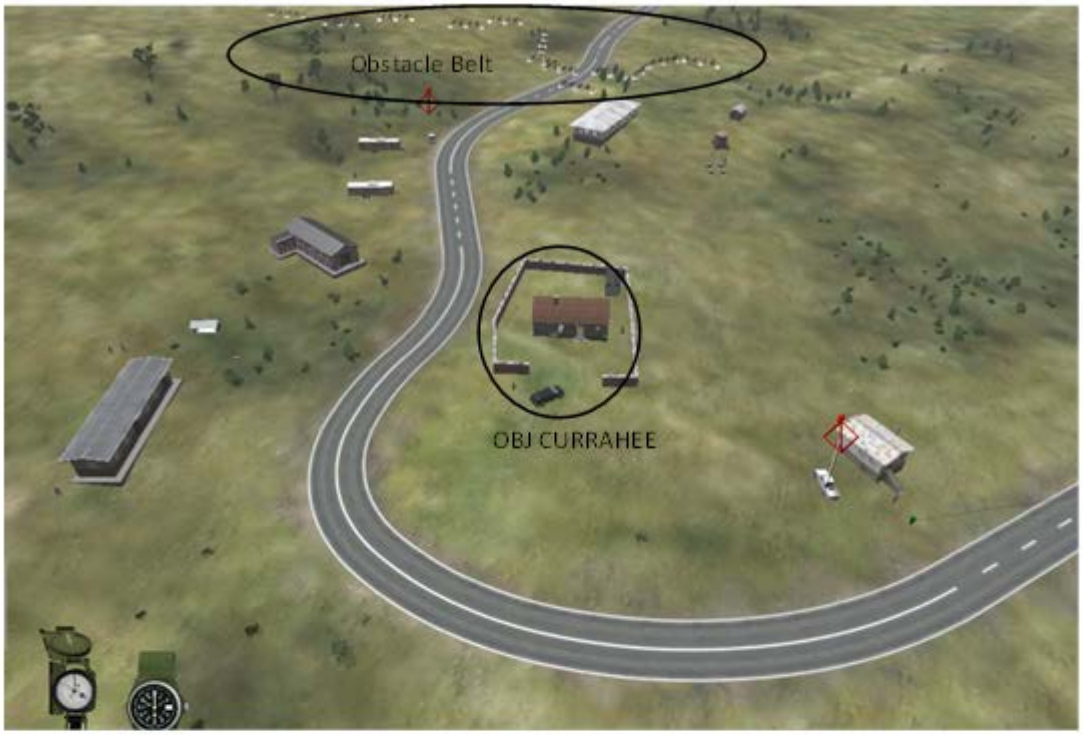

CONTOUR INTERVAL 100 m
Elevations in meters
200 m
OBJ CURRAHEE
Oak Grove Airfield
Oak Grove
Trenton
N


CONTOUR INTERVAL 10 m
Elevations in meters
20 m
OBJ CURRAHEE
Obstacle Belt
N


R
P
N


OBJ CURRAHEE
Obstacle Belt
N


OBJ CURRAHEE
Obstacle Belt
N


Obstacle Belt
OBJ CURRAHEE


Obstacle Belt
OBJ CURRAHEE

LIST OF REFERENCES

Andersen, E., Liu, Y. E., Snider, R., Szeto, R., Cooper, S., & Popović, Z. (2011, June). On the harmfulness of secondary game objectives. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, 30–37.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving an introduction and cases. *Convergence: The International Journal Of Research into New Media Technologies*, *14*(1), 75–90.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization*, *90*, 123–133.

Department of Defense (2005). *Joint capabilities integration and development system* (Chairman of the Joint Chiefs of Staff Instruction CJCSI 3170.01E). Washington, D.C.: Author.

Department of Defense. (2009). *Incorporating test and evaluation into Department of Defense acquisition contracts.* Washington, D.C.: Author.

Department of Defense (2015). *Operation of the Defense acquisition system* Department of Defense Instruction 5000.02). Washington, D.C.: Author.

Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. *Communications of the ACM*, *54*(4), 86–96.

El-Nasr, M. S., Drachen, A., & Canossa, A. (2013). *Game analytics: Maximizing the value of player data*. Berlin, Germany: Springer Science & Business Media.

Fitz-Walter, Z. (2013). A brief history of gamification. Retrieved from http://zefcan.com/2013/01/a-brief-history-of-gamification

Freedberg, S. J. (2014, August 6). We've got to wake up: Frank Kendall calls for defense innovation. Retrieved from http://breakingdefense.com/2014/08/wevegot-to-wake-up-frank-kendall-calls-for-defense-innovation

Goh, D. H., & Lee, C. S. (2011). Perceptions, quality and motivational needs in image tagging human computation games. *Journal of Information Science*, doi: 10.1177/0165551511417786

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, *14*(6), 1–4.

McGroarty, C. *Innovation and rapid evolutionary design by virtual doing: understanding Early Synthetic Prototyping (ESP).* Retrieved from http://ict.usc.edu/pubs/Innovation%20and%20Rapid%20Evolutionary%20Design

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013, April). Disassembling gamification: the effects of points and meaning on user motivation and performance. In *CHI'13 Extended Abstracts On Human Factors In Computing Systems*, 1137–1142.

Murray, K. L. (2014). *Early synthetic prototyping: Exploring designs and concepts within games*. Naval Postgraduate School, Monterey CA.

Parker, P. E. (2014, September 3). In Newport, Hagel says defense establishment must push for greater innovation. *Providence Journal*. Retrieved from http://www.providencejournal.com/breaking-news/content/20140903-innewport-hagel-says-defense-establishment-must-push-for-greater-innovation.ece

Perkins, T., Peterson, R., & Smith, L. (2003, December). Back to the Basics: Measurement and Metrics. *The Journal of Defense Software Engineering*, 9–12.

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas? Journal of Product Innovation Management, *29*(2), 245–256.

Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor.

U.S. Army (2014). *Army training and leader development* (Army Regulation 350-1). Washington, D.C.: Author.

U.S. Army. (2011). *System training integration* (TRADOC Pamphlet 350-70-13). Washington, D.C.: Author.

Vogt, B. D., Megiveron, M. G., & Smith, R. E. (2015). Early synthetic prototyping: When we build it, will they come? *I/ITSEC Proceedings, 2014*, Orlando, FL.

Wang, H., & Sun, C. T. (2011, September). Game reward systems: gaming experiences and social meanings. In *Proceedings of DiGRA 2011 Conference: Think Design Play*, 1–12.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California